

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 6, 2002, 09:38:31 ; Search time 12.18 Seconds
(without alignments)
29.561 Million cell updates/sec

Title: US-09-897-042-18

Sequence: 1 MNLVPMNPLVNMFEFM 16

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 96934

Minimum DB seq length: 0
Maximum DB seq length: 16

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 50 summaries

Database : Issued Patents-AA:*
1: /cgn2_6/prodata/2/1aa/5A_COMB.pep:*
2: /cgn2_6/prodata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/prodata/2/1aa/6A_COMB.pep:*
4: /cgn2_6/prodata/2/1aa/6B_COMB.pep:*
5: /cgn2_6/prodata/2/1aa/PCTUS_COMB.pep:*
6: /cgn2_6/prodata/2/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	27	31.8	10	2	US-07-963-538B-14
2	27	31.8	11	3	US-08-951-944-8
3	27	31.8	15	4	US-08-408-930A-4
4	26	30.6	9	3	US-09-075-257A-1
5	26	30.6	9	4	US-09-534-639-1
6	26	30.6	10	3	US-09-075-257A-10
7	26	30.6	10	4	US-09-534-639-10
8	26	30.6	13	1	US-08-128-971B-2
9	26	30.6	15	4	US-08-256-104-5
10	26	30.6	16	4	US-08-602-999A-172
11	25	29.4	14	1	US-08-160-317-4
12	25	29.4	14	1	US-08-351-147-4
13	25	29.4	14	1	US-08-471-154-4
14	25	29.4	16	1	US-08-574-763-7
15	25	29.4	16	4	US-08-602-999A-234
16	25	29.4	16	5	PCT-US91-09422-31
17	24.5	28.8	12	3	US-08-951-944-10
18	24	28.2	9	4	US-09-177-249-299
19	24	28.2	10	3	US-08-159-339A-474
20	24	28.2	10	3	US-08-159-339A-475
21	24	28.2	10	4	US-08-685-558A-2
22	24	28.2	13	1	US-08-464-531-106
23	24	28.2	13	2	US-08-461-598-106
24	24	28.2	13	3	US-08-332-137-106
25	24	28.2	13	4	US-08-582-333A-42
26	24	28.2	14	1	US-07-991-199D-5
27	24	28.2	14	3	US-09-082-039A-12

ALIGNMENTS

28	24	28.2	14	5	PCT-US93-12246-5	Sequence 5, Appl
29	24	28.2	15	1	US-08-221-583-61	Sequence 61, Appl
30	24	28.2	15	1	US-08-221-583-62	Sequence 62, Appl
31	24	28.2	15	1	US-07-965-667A-19	Sequence 19, Appl
32	24	28.2	15	3	US-08-484-819-19	Sequence 19, Appl
33	24	28.2	15	4	US-08-602-999A-171	Sequence 171, App
34	24	28.2	15	5	PCT-US93-10197-19	Sequence 19, Appl
35	24	28.2	15	5	PCT-US95-04018-61	Sequence 61, Appl
36	24	28.2	15	5	PCT-US95-04018-62	Sequence 62, Appl
37	23	27.1	7	2	US-08-632-598-11	Sequence 11, Appl
38	23	27.1	7	4	US-09-231-240-11	Sequence 11, Appl
39	23	27.1	8	1	US-08-636-386-4	Sequence 4, Appl
40	23	27.1	8	5	PCT-US95-13846-4	Sequence 4, Appl
41	23	27.1	9	2	US-08-318-856A-53	Sequence 53, Appl
42	23	27.1	9	4	US-09-518-046-110	Sequence 110, App
43	23	27.1	10	1	US-08-636-386-5	Sequence 5, Appl
44	23	27.1	10	1	US-08-408-604A-71	Sequence 71, Appl
45	23	27.1	10	5	PCT-US95-13846-5	Sequence 5, Appl
46	23	27.1	13	1	US-08-189-772-2	Sequence 2, Appl
47	23	27.1	13	1	US-08-188-277B-11	Sequence 11, Appl
48	23	27.1	13	2	US-08-467-046-7	Sequence 7, Appl
49	23	27.1	14	1	US-08-188-277B-24	Sequence 24, Appl
50	23	27.1	15	1	US-08-449-207-4	Sequence 4, Appl

RESULT 1
US-07-963-538B-14
; Sequence 14, Application US/07963538B
; Patent No. 5851983
; GENERAL INFORMATION:
; APPLICANT: SUGIYAMA, TAKASHI
; APPLICANT: KAMIMURA, TAKASHI
; APPLICANT: MASUDA, KENICHI
; APPLICANT: OKADA, MASAHIRO
; APPLICANT: OHTSUKA, EIKO
; APPLICANT: IMAIZUMI, ATSUSHI
; APPLICANT: MATSUMOTO, YOHICHI
; APPLICANT: MATSUMOTO, YOHICHI
; APPLICANT: TAKEUCHI, AKIKO
; TITLE OF INVENTION: ELASTASE INHIBITORY POLYPEPTIDE AND
; TITLE OF INVENTION: PROCESS FOR PRODUCTION THEREOF BY RECOMBINANT GENE
; TITLE OF INVENTION: TECHNOLOGY
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: COOLEY GODWARD LLP
; STREET: FIVE PALO ALTO SQUARE, 4TH FLOOR
; CITY: PALO ALTO
; STATE: CA
; COUNTRY: USA
; ZIP: 94306-2155
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/963,538B
; FILING DATE: 20-OCT-1992
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/843,359
; FILING DATE: 25-FEB-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/408,483
; FILING DATE: 22-AUG-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 4-212399
; FILING DATE: 17-JUL-1992
; PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 4-212398
FILING DATE: 17-JUL-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 3-355553
FILING DATE: 24-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 62-330219
FILING DATE: 28-DEC-1987
ATTORNEY/AGENT INFORMATION:
NAME: NEELEY PH.D., RICHARD L.
REGISTRATION NUMBER: 30,092
REFERENCE/DOCKET NUMBER: TEJN-005/0205
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-843-5070
TELEFAX: 415-857-0663
TELEX: 380816 COOLEY PA
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 10 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-07-963-538B-14

Query Match 31.8%; Score 27; DB 2; Length 10;
Best Local Similarity 83.3%; Pred. No. 67;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 LVPMNP 8
11111
DB 4 LVPMNP 9

RESULT 2
US-08-951-944-8
Sequence 8, Application US/08951944
Patent No. 6037449
GENERAL INFORMATION:
APPLICANT: Zhou, Hai-Yan
APPLICANT: Salih, Erdan
APPLICANT: Glincher, Melvin J.
TITLE OF INVENTION: OSTEOETRIN RELATED PEPTIDES AND
TITLE OF INVENTION: NUCLEIC ACIDS AND DIAGNOSIS AND THERAPIES USING THEM
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: FastSeq for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/951,944
FILING DATE: 17-OCT-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/028,682
FILING DATE: 18-OCT-1996
ATTORNEY/AGENT INFORMATION:
NAME: Freeman, John W.
REGISTRATION NUMBER: 29,066
REFERENCE/DOCKET NUMBER: 00108/125001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:

LENGTH: 11 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-951-944-8

Query Match 31.8%; Score 27; DB 3; Length 11;
Best Local Similarity 62.5%; Pred. No. 75;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 PMNPVNM 12
11111
DB 3 PMNPVNM 10

RESULT 3
US-08-408-930A-4
Sequence 4, Application US/08408930A
Patent No. 6242567
GENERAL INFORMATION:

APPLICANT: Hema Pandey, Arthur D. Riggs, John
APPLICANT: A. Zaia and Brian R. Clark
TITLE OF INVENTION: Method For Detection and
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: City of Hope
STREET: 1500 East Duarte Road
CITY: Duarte
STATE: California
COUNTRY: United States of America
ZIP: 91010-0269

COMPUTER READABLE FORM:
MEDIUM TYPE: 3M Double Density
MEDIUM TYPE: 5 1/4" diskette
COMPUTER: Wang PC
OPERATING SYSTEM: MS DOS Version 3.20
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/408,930A
FILING DATE: 22 March 1995

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/978,151
FILING DATE: 17 No. 6242567ember 1992
ATTORNEY/AGENT INFORMATION:
NAME: Irons, Edward S.
REGISTRATION NUMBER: 16,541
REFERENCE/DOCKET NUMBER: Pandey
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 626-3564
TELEFAX: (202) 783-6031
TELEX: No. 6242567e

INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 15
TYPE: Amino Acid
STRANDEDNESS: Single
TOPOLOGY: Unknown
US-08-408-930A-4

Query Match 31.8%; Score 27; DB 4; Length 15;
Best Local Similarity 50.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 2 NLVPMNPVNM 13
1111111
DB 1 NLVPMNPVNM 12

RESULT 4

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 6, 2002, 09:36:16 ; Search time 25.16 Seconds
(Without alignments)
47.105 Million cell updates/sec

Title: US-09-897-042-18
Perfect score: 85
Sequence: 1 MNLVPMNPLVNMEEFN 16

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues
Total number of hits satisfying chosen parameters: 178144

Minimum DB seq length: 0
Minimum DB seq length: 16

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database :

A.Geneseq_1101.*
1: /SID52/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	31.8	15	17	AA097927	Japan cedar pollen
2	32.9	11	19	AA083283	NPF motif EH domain
3	31.8	9	17	AA000715	Human Stat 5 fragm
4	31.8	9	22	AA007384	HLA-A1 nonamer #1
5	31.8	9	22	AA007797	HLA-A *0201 noname
6	31.8	9	22	AA008475	HLA-B *2705 noname
7	31.8	9	22	AA009077	HLA-B *0702 noname
8	31.8	9	22	AA011238	HLA-A26 nonamer #1
9	31.8	9	22	AA012037	HLA-B8 nonamer #27
10	31.8	9	22	AA012642	HLA-B *1510 noname
11	31.8	10	22	AA009183	HLA-A *0201 decame

12	27	31.8	10	22	AA009603	HLA-A1 decamer #17
13	27	31.8	10	22	AA009633	HLA-A1 decamer #20
14	27	31.8	10	22	AA009647	HLA-B *0702 decame
15	27	31.8	10	22	AA010021	HLA-B *0702 decame
16	27	31.8	10	22	AA010974	HLA-A *0201 decame
17	27	31.8	10	22	AA012823	HLA-A26 decamer #6
18	27	31.8	10	22	AA013449	HLA-A26 decamer #5
19	27	31.8	14	22	AA098147	Human SNP associat
20	27	31.8	15	19	AA062870	Epitope of the p66
21	27	31.8	15	20	AA031203	Human cytomegalov
22	27	31.8	15	22	AA084955	HCMV 64 kilodalton
23	26	30.6	8	22	AA009352	HLA-B8 octamer #35
24	26	30.6	8	22	AA009520	HLA-B8 octamer #79
25	26	30.6	8	22	AA010247	HLA-B8 octamer #26
26	26	30.6	8	22	AA010943	HLA-B8 octamer #42
27	26	30.6	9	19	AA062869	Epitope of the p66
28	26	30.6	9	19	AA048979	Nonpeptide immuno
29	26	30.6	9	20	AA031196	Human cytomegalov
30	26	30.6	9	20	AA031197	Human cytomegalov
31	26	30.6	9	20	AA031198	Human cytomegalov
32	26	30.6	9	20	AA009319	Immunogenic peptid
33	26	30.6	9	21	AA012406	Immunogenic epitop
34	26	30.6	9	21	AA054994	HLA-A2 restricted
35	26	30.6	9	22	AA007208	HLA-B *0702 noname
36	26	30.6	9	22	AA007232	HLA-B *0702 noname
37	26	30.6	9	22	AA007249	HLA-B *0702 noname
38	26	30.6	9	22	AA007321	HLA-B *0201 noname
39	26	30.6	9	22	AA007427	HLA-A1 nonamer #13
40	26	30.6	9	22	AA007632	HLA-A1 nonamer #16
41	26	30.6	9	22	AA007749	HLA-B *2705 noname
42	26	30.6	9	22	AA007906	HLA-B *2705 noname
43	26	30.6	9	22	AA008008	HLA-B *2705 noname
44	26	30.6	9	22	AA008057	HLA-A *0201 noname
45	26	30.6	9	22	AA008264	HLA-A *0201 noname
46	26	30.6	9	22	AA008476	HLA-B *2705 noname
47	26	30.6	9	22	AA008532	HLA-A *0201 noname
48	26	30.6	9	22	AA008714	HLA-B *2705 noname
49	26	30.6	9	22	AA008820	HLA-A1 nonamer #46
50	26	30.6	9	22	AA008821	HLA-A1 nonamer #46

ALIGNMENTS

RESULT 1
AA097927 ID AA097927 standard; peptide: 15 AA.
XX
AC AA097927;
XX
DT 16-AUG-1996 (first entry)
XX
DE Japan cedar pollen mature allergen Cry j II amino acids 281-295.
XX
KW Allergen; epitope; overlapping peptide; Cry j II; cedar pollen;
KW Sugi pollinosis; diagnosis; treatment.
XX
OS Cryptomeria japonica.
XX
FN JP08047392-A.
XX
PD 20-FEB-1996.
XX
PE 07-NOV-1994; 94JP-0297840.
XX
PR 26-MAY-1994; 94JP-0134868.
XX
PR 05-NOV-1993; 93JP-0276773.
XX
PA (MEIP) MEIJI MILK PROD CO LTD.
XX
DR WPI: 1996-166249/17.
XX
PT Japan cedar pollen allergen Cry j II epitope - comprises at least

PT part of specified 460 amino acid protein
XX
PS Disclosure; Fig 5; 17pp; Japanese.
XX
CC AAR97871-R97960 are overlapping peptides used for the epitope mapping
CC of the Japan cedar pollen allergen Cry j II. Cry j II and allergenic
CC peptides of it are useful in the diagnosis, prevention and treatment
CC of Sugi pollinosis, the allergic reaction to Japan cedar pollen.
CC Significant regions of the allergen were identified using the
CC overlapping peptides of the full epitope derived from a Cry j II
CC antigen-specific T cell line. Amino acids 66-80 (AAR97884) and 186-200
CC (AR978808) of the full mature 460 amino acid allergen are the most
CC allergenic of the 90 peptides tested.
XX
SQ Sequence 15 AA:

Query Match 36.5%; Score 31; DB 17; Length 15;
Best Local Similarity 50.0%; Pred. No. 41;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 7 NPLVNER 14
|:|:|:|:|:
Db 4 npllnqf 11

RESULT 2
AAW83283
ID AAW83283 standard; peptide: 11 AA.
XX
AC AAW83283;
XX
DT 08-FEB-1999 (first entry)
XX
DE NPF motif EH domain binding peptide #64.
XX
KW Human: cytoplasmic protein; EH-containing protein; eps15; eps15R;
KW eps15 homology; intracellular interaction; EH domain binding specificity;
KW signal transducer; NPF motif; h-NUMB; h-NUMB-R; h-RAB; h-RAB-R; ehb3;
KW ehb10; ehb21; cell proliferation; diagnosis; detection.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO9846744-A1.
XX
DD 22-OCT-1998.
XX
PF 06-APR-1998; 98WO-IT00077.
XX
PR 15-APR-1997; 97IT-MI00868.
XX
PA (EUON-) IST EURO DI ONCOLOGIA SRL.
XX
PI Di Fiore PP, Doria M, Pelicci PG, Salcini AE;
XX
DR WPI: 1998-594574/50.
XX
PT New isolated EH domain binding proteins and peptide(s) - obtained
PT using signal transducers eps15 and eps15R containing EH domains by
PT detecting specific binding activity.
XX
PS Claim 11; Page 73; 90pp; English.
XX
CC The present invention describes a new intracellular interactor and novel
CC protein:protein interaction (EH) domain binding protein having (parts
CC of) one of the following human derived sequences designated: (I) h-NUMB;
CC (II) h-NUMB-R; (III) h-RAB-R; (IV) ehb3; (V) ehb10; or (VI) ehb21. The
CC present invention also describes peptides containing at least one NPF
CC (ASP-Pro-Phe) motif, able to bind to a protein with at least one EH
CC domain. The proteins with a NPF-containing peptide or the peptides
CC themselves can be used to identify and purify EH containing proteins.
CC sense RNA, complementary to mRNA encoding h-NUMB, h-NUMB-R, h-RAB-R,

CC ehb3, ehb10, or ehb21, can be used for diagnostic and therapeutic uses.
CC products from the present invention can also be used to develop agents
CC for use in control of cell proliferation. AAW83220 to AAW83284 represent
CC NPF motif EH domain binding peptides.
XX
SQ Sequence 11 AA:

Query Match 32.9%; Score 28; DB 19; Length 11;
Best Local Similarity 44.4%; Pred. No. 97;
Matches 4; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 PMNPVME 13
|:|:|:|:|:
Db 3 pmnpfvmq 11

RESULT 3
AAW00715
ID AAW00715 standard; Peptide: 9 AA.
XX
AC AAW00715;
XX
DT 06-DEC-1996 (first entry)
XX
DE Human Stat 5 fragment (aa735-743).
XX
KW Signal transducer and activator of transcription 5; Stat 5;
KW Interleukin-2; signal transduction; cell proliferation;
KW immune disorder; therapy.
XX
OS Homo sapiens.
XX
PN WO9626292-A1.
XX
PD 29-AUG-1996.
XX
PF 22-FEB-1996; 96WO-US02457.
XX
PR 23-FEB-1995; 95US-0393333.
XX
PA (TULDA-) TULARIK INC.
XX
PI Hou J, McKnight SL, Schindler U;
XX
DR WPI: 1996-402382/40.
XX
PT Human signal transducer and activator of transcription 5 protein -
PT used for treating cellular proliferation disorders, pref. of immune
PT cells
XX
PS Claim 2; Page 36; 42pp; English.
XX
CC Fragments (AAW00707-20) of human signal transducer and activator of
CC transcription 5 (hstat 5) (AAW00706) retain hstat 5-specific
CC binding affinity and are capable of eliciting antibodies capable
CC of distinguishing hstat 5 from other Stats and Stat-related
CC proteins such as mammary gland factor. They can be produced
CC by chemical synthesis or by recombinant methods. They are
CC useful e.g. in two-hybrid screening to identify intracellular
CC targets which specifically bind to hstat 5. Screening assays
CC incorporating the fragments are used to identify agents useful
CC in the diagnosis, prognosis or treatment of diseases associated
CC with undesirable cell growth, differentiation and/or cytokine
CC signal responsiveness.
XX
SQ Sequence 9 AA:

Query Match 31.8%; Score 27; DB 17; Length 9;
Best Local Similarity 57.1%; Pred. No. 4.3e+05;
Matches 4; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
 DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
 DE NEURAL CELL ADHESION MOLECULE (FRAGMENT).
 GN NCAM.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BALB-C; TISSUE=LIVER;
 RX MEDLINE=98250618; PubMed=9582442;
 RA Kawahigashi H., Harada Y., Asano A., Nakamura M.;
 RT "A cis-acting regulatory element that affects the alternative splicing
 of a muscle-specific exon in the mouse NCAM gene."
 RL Blochum. Biophys. Acta 1397:305-315(1998).
 DR EMBL: AB001873; BAA31274.1; -.
 DR MGD; MGI:97281; Ncam.
 FT NON_TER 1
 SQ SEQUENCE 15 AA: 1481 MW: 41868EP61I7732C2 CRC64;

Query Match 29.4%; Score 25; DB 11; Length 15;
 Best Local Similarity 66.7%; Pred. No. 8.8e+02;
 Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 3 LVPKMP 8
 |||:|
 DB 8 LVPKMP 13

RESULT 3
 ID O9PTM6 PRELIMINARY; PRT: 15 AA.
 AC O9PTM6;
 DT 01-MAY-2000 (TReMBLrel. 13, Created)
 DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
 DE 01-MAY-2001 (TReMBLrel. 16, Last annotation update)
 GN AROMATASE (FRAGMENT).
 OS AROMATASE.
 OS Pterophila guttata (zebra finch) (Taeniopygia guttata).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Passeriformes; Estrilidae;
 OC Estrilidae; Taeniopygia.
 OC NCBI_TaxID=59729;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20033564; PubMed=10564828;
 RA Ramachandran B., Schlinger B.A., Arnold A.P., Campagnoni A.T.;
 RT "zebra finch aromatase gene expression is regulated in the brain
 through an alternate promoter."
 RL Gene 240:209-216(1999).
 DR EMBL: AF170273; AAF20041.1; -.
 FT NON_TER 15
 SQ SEQUENCE 15 AA: 1745 MW: 29099C05I46F360E CRC64;

Query Match 28.2%; Score 24; DB 13; Length 15;
 Best Local Similarity 50.0%; Pred. No. 1.3e+03;
 Matches 6; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

OY 1 MNLVPMNPLVMN 12
 | | : | | |
 DB 1 MVLTLNPLVHN 12

RESULT 4
 ID O36789 PRELIMINARY; PRT: 16 AA.
 AC O36789;
 DT 01-NOV-1996 (TReMBLrel. 01, Created)
 DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)

DT 01-MAY-1999 (TReMBLrel. 10, Last annotation update)
 DE PPS19' PROTEIN (FRAGMENT).
 GN PPS19'.
 OS Solanum nigrum (Black nightshade).
 OC Chloroplast.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots;
 OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
 OC NCBI_TaxID=4112;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96397499; PubMed=8804393;
 RA Goulding S.E., Olmstead R.G., Morden C.W., Wolfe K.H.;
 RT "Bb and flow of the chloroplast inverted repeat."
 RL Mol. Gen. Genet. 252:195-206(1996).
 DR EMBL: Z71249; CAA94964.1; -.
 KW Chloroplast.
 FT NON_TER 16
 SQ SEQUENCE 16 AA: 1869 MW: 1AB9DE1D775A138C CRC64;

Query Match 28.2%; Score 24; DB 8; Length 16;
 Best Local Similarity 44.4%; Pred. No. 1.4e+03;
 Matches 4; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 7 NPLVNMFM 15
 || | | :
 DB 8 NPVAVNHL 16

RESULT 5
 ID O54325 PRELIMINARY; PRT: 15 AA.
 AC O54325;
 DT 01-NOV-1996 (TReMBLrel. 01, Created)
 DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
 DT 01-NOV-1998 (TReMBLrel. 08, Last annotation update)
 DE CHLORAMPHENICOL ACETYLTRANSFERASE (FRAGMENT).
 GN CAT.
 OS Staphylococcus intermedius.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Staphylococcus.
 OC NCBI_TaxID=1285;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-PCM6;
 RX MEDLINE=96379895; PubMed=8787908;
 RA Lodder G., Schwarz S., Gregory P., Dyke K.;
 RT "Tandem duplication in ermC translational attenuator of the macrolide-
 R lincosamide-streptogramin B resistance plasmid pSES6 from
 RT Staphylococcus equorum."
 RL Antimicrob. Agents Chemother. 40:215-217(1996).
 DR EMBL: X82666; CAA57981.1; -.
 KW Transferase.
 FT NON_TER 1
 SQ SEQUENCE 15 AA: 1931 MW: 173A6AA93999A8D9 CRC64;

Query Match 27.1%; Score 23; DB 2; Length 15;
 Best Local Similarity 66.7%; Pred. No. 1.9e+03;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 11 MNEFMN 16
 || | | :
 DB 1 MNERFD 6

RESULT 6
 ID P93232 PRELIMINARY; PRT: 16 AA.
 AC P93232;
 DT 01-MAY-1997 (TReMBLrel. 03, Created)
 DT 01-MAY-1997 (TReMBLrel. 03, Last sequence update)

GenCore version 4.5.
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 6, 2002, 09:40:42 ; Search time 9.79 Seconds
(without alignments)
59.922 Million cell updates/sec

Title: US-09-897-042-18

Perfect score: 85
Sequence: 1 MNLVPMNPLVMNEFMN 16

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 780

Minimum DB seq length: 0
Maximum DB seq length: 16

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	28	32.9	15	1	UBI1_MONDO
2	22.5	26.5	15	1	PC20_BRANA
3	22	25.9	15	1	TM2A_METMA
4	20.5	24.1	15	1	UC29_MAIZE
5	20	23.5	9	1	UPA7_HUMAN
6	19	22.4	8	1	FAR8_CALYO
7	19	22.4	11	1	FAR8_CALYO
8	19	22.4	11	1	FAR8_CALYO
9	18.5	21.8	14	1	ATP6_SPIOL
10	18.5	21.8	14	1	ATP6_SPIOL
11	18	21.2	13	1	PHGR_EUBOX
12	18	21.2	11	1	TKN3_PSEGU
13	18	21.2	11	1	TKN3_PSEGU
14	18	21.2	13	1	CRBL_VESCR
15	18	21.2	13	1	CRBL_VESCR
16	18	21.2	14	1	CRBL_VESCR
17	18	21.2	14	1	CRBL_VESCR
18	18	21.2	15	1	CRBL_VESCR
19	18	21.2	15	1	CRBL_VESCR
20	18	21.2	15	1	CRBL_VESCR
21	18	21.2	16	1	CRBL_VESCR
22	17	20.0	6	1	VP19_HSV1K
23	17	20.0	11	1	CS15_BACSU
24	17	20.0	11	1	CS15_BACSU
25	17	20.0	11	1	CS15_BACSU
26	17	20.0	15	1	RBS_PHRPA
27	17	20.0	15	1	RBS_PHRPA
28	17	20.0	15	1	RBS_PHRPA
29	17	20.0	15	1	RBS_PHRPA
30	16.5	19.4	13	1	CRBL_ICASP
31	16	18.8	9	1	CRBL_ICASP
32	16	18.8	10	1	RRPL_PHODV
33	16	18.8	10	1	RRPL_PHODV

ALIGNMENTS

RESULT	1	STANDARD	PRT	15 AA.
UBI1_MONDO				
ID	UBI1_MONDO			
AC	P50103			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	20-AUG-2001 (Rel. 40, Last annotation update)			
DE	UBIQUITIN CARBOXYL-TERMINAL HYDROLASE ISOZYME L1 (EC 3.4.19.12) (UCH-L1) (UBIQUITIN THIOLESTERASE L1) (Pgp 9.5) (FRAGMENT).			
DE	UCHL1.			
OS	Monodelphis domestica (Short-tailed grey opossum).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Metatheria; Didelphimorphia; Didelphidae; Monodelphis.			
OX	NCBI_TaxID=13616;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Liver;			
RX	MEDLINE=96102916; PubMed=8522974;			
RA	Mann D.A., Trowern A.R., Lavender F.L., Wittaker P.A.,			
RA	Thompson R.J.;			
RT	"Identification of evolutionary conserved regulatory sequences in the			
RT	5' untranslated region of the neural-specific ubiquitin C-terminal			
RT	hydrolyase (Pgp9.5) gene."			
RL	J. Neurochem. 66:35-46(1996).			
CC	-1- FUNCTION: UBIQUITIN-PROTEIN HYDROLASE IS INVOLVED BOTH IN THE			
CC	PROCESSING OF UBIQUITIN PRECURSORS AND OF UBIQUITINATED PROTEINS.			
CC	A PEPTIDE BOND AT THE C-TERMINAL GLYCINE OF UBIQUITIN.			
CC	-1- CATALYTIC ACTIVITY: UBIQUITIN C-TERMINAL THIOLESTER + H(2)O =			
CC	UBIQUITIN + A THIOL.			
CC	-1- SUBCELLULAR LOCATION: CYTOPLASMIC.			
CC	-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C12; ALSO KNOWN AS			
CC	FAMILY 1 OF UBIQUITIN CARBOXYL-TERMINAL HYDROLASES.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@lsb-sib.ch).			
CC	-----			
DR	EMBL: U32208; AAA8059.1; -			
DR	InterPro: IPR001578; UCH.			
DR	PROSITE: PS00140; UCH_1; PARTIAL.			
KW	Ubiquitin conjugation; Hydrolyase; Thiol protease; Multi-gene family.			
FT	NON_TER 15			
SO	SEQUENCE 15 AA; 1816 MW; 1B7A3B6E23F6E79 CRC64;			

Query Match

32.9%; Score 28; DB 1; Length 15;

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 6, 2002, 09:38:06 ; Search time 19.06 Seconds
(without alignments)
63.945 Million cell updates/sec

Title: us-09-897-042-18

Perfect score: 85

Sequence: 1 MNLVPMNPLVMEFMN 16

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 2758
Minimum DB seq length: 0
Maximum DB seq length: 16

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database :
1: PIR68:*
2: PIR1:*
3: PIR2:*
4: PIR3:*
5: PIR4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	35	41.2	16	2 A45454	ankyrin-binding gl
2	28	32.9	15	2 S10891	ubiquitin thiole
3	28	32.9	15	2 P70093	ubiquitin thiole
4	23	27.1	9	2 A60356	118k stomach cance
5	22	25.9	11	2 S09024	carboxylesterase (
6	21	24.7	11	2 A48973	glucosylase A1 (E
7	21	24.7	12	2 S29859	gene p10 protein -
8	20	23.5	9	2 P70080	60k Ca binding pro
9	20	23.5	12	2 P70080	ATP synthase D cha
10	20	23.5	14	2 B61309	lutropin beta chal
11	20	23.5	14	2 F83754	hypothetical prote
12	20	23.5	15	2 PA0060	protein QP200037 -
13	20	23.5	15	2 B61457	alpha-glucosidase
14	20	23.5	16	2 A24099	crystal protein, 2
15	19	22.4	5	2 B37325	pap fibriaral regul
16	19	22.4	8	2 A39308	glycine reductase
17	19	22.4	8	2 H41978	callifmrfamide 8 -
18	19	22.4	11	2 S14196	3-hydroxy-3-methyl
19	19	22.4	12	2 S74196	H+-transporting AT
20	19	22.4	14	2 S21247	collagen alpha cha
21	19	22.4	14	2 S23376	H+-transporting AT
22	19	22.4	14	2 G33160	pseudogerm - whe
23	19	22.4	15	2 P00545	protein QP100022 -
24	19	22.4	15	2 S27248	protein QP200007 -
25	19	22.4	15	2 PA0090	protein QP200011 -
26	19	22.4	15	2 PA0092	adherence lectin h
27	19	22.4	15	2 A31435	heat shock protein
28	19	22.4	15	2 S71306	

30	19	22.4	15	2 B59137	protein Pfl - gold
31	19	22.4	16	2 I57530	gene c-fms protein
32	19	22.4	16	2 S11760	flgl protein - Cau
33	19	22.4	16	2 PH1302	Ig heavy chain DJ
34	19	22.4	16	2 S09084	protease chain 3
35	18.5	21.8	11	2 F60409	substance P-like p
36	18.5	21.8	13	2 A32453	phloroglucinol red
37	18.5	21.8	13	2 B24099	crystal protein, 7
38	18.5	21.8	16	2 A44692	fullcin - giant Af
39	18	21.2	5	2 A44692	granulocyte-colony
40	18	21.2	8	2 I54017	kassinin-like pept
41	18	21.2	11	2 D60409	cycloimmunoligosac
42	18	21.2	11	2 PC2330	T antigen variant
43	18	21.2	11	2 PH1375	cytochrome P450 3A
44	18	21.2	12	2 PC2254	exo-alpha-sialidas
45	18	21.2	12	2 S11286	crabrolin - Europe
46	18	21.2	13	2 JZVHP1	seed protein ws-4
47	18	21.2	13	2 D61491	cell surface glyco
48	18	21.2	13	2 A40207	locustamytroptin I
49	18	21.2	13	2 B61620	Ig kappa chain V-I
50	18	21.2	13	2 E61458	deoxynucleotidyltr

ALIGNMENTS

RESULT 1

A45454
ankyrin-binding glycoprotein ABGP, 186k (N-terminal) - rat (fragment)

C:Species: Rattus norvegicus (Norway rat)
C:Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 07-Feb-1997

C:Accession: A45454
R: Davis, J.Q.; McLaughlin, T.; Bennett, V.

J. Cell Biol. 121, 121-133, 1993

A:Title: Ankyrin-binding proteins related to nervous system cell adhesion molecules:
A:Reference number: A45454; MUID:93209979

A:Accession: A45454
A>Status: preliminary

A:Molecule type: protein
A:Residues: 1-16 <DAV>

A:Experimental source: brain
A:Note: Sequence extracted from NCBI backbone (NCBI:P128327)

C:Keywords: glycoprotein

Query Match	41.2%	Score 35;	DB 2;	Length 16;
Best Local Similarity	50.0%	Pred. NO. 6.1;		
Matches	5;	Conservative	3;	Mismatches
			2;	Indels
				Gaps
				0;

RESULT 2

S10891
ubiquitin thiolesterase (EC 3.1.2.15) PGP9.5 - human (fragment)

N:Alternate names: ubiquitin carboxyl-terminal hydrolase; neuron-specific

C:Species: Homo sapiens (man)
C:Date: 20-Feb-1995 #sequence_revision 29-May-1998 #text_change 11-Jan-2000

C:Accession: S10891
R: Day, I.N.M.; Hinks, L.J.; Thompson, R.J.

Biochem. J. 268, 521-524, 1990
A:Title: The structure of the human gene encoding protein gene product 9.5 (PGP9.5),
A:Reference number: S10891; MUID:90303237

A:Accession: S10891
A:Molecule type: DNA

A:Residues: 1-15 <DAV>
A:Cross-references: EMBL:X17377; NID:g35441; PIDN:CAA35249.1; PID:g296799

C:Genetics: 11/3

C:Superfamily: human ubiquitin thiolesterase
C:Keywords: thiolester hydrolase

Query Match 32.9%; Score 28; DB 2; Length 15;
Best Local Similarity 46.7%; Pred. No. 89;
Matches 7; Conservative 3; Mismatches 3; Indels 2; Gaps 1;

OY 1 MNLVPM--NPLVME 13
| | | | |
Db 1 MQLKPMENPEMLNK 15

RESULT 3
PT0093
ubiquitin thiolesterase (EC 3.1.2.15) - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 21-Aug-1998 #sequence_revision 21-Aug-1998 #text_change 18-Sep-1998
C:Accession: PT0093; PN0045

R:Kawakami, T.; Uchida, T.; Sakai, T.; Kamo, M.; Morimasa, T.; Tsugita, A.
C:Title: Analysis of proteins isolated by two dimensional electrophoresis of mouse neur
C:Comment: The molecular mass is 28,000 and the pI is 5.0.
C:Keywords: brain; thiolester hydrolase

A:Reference number: PT0091
A:Accession: PT0093

A:Molecule type: protein
A:Residues: 1-15 <KAM>
A:Experimental source: brain, striatum

R:Kato, H.

Kawasaki Igakkaishi 22, 245-259, 1996
A:Title: Analysis of proteins isolated by two dimensional electrophoresis of mouse neur

A:Reference number: PN0041

A:Accession: PN0045

A:Molecule type: protein

A:Residues: 1-15 <KAT>

A:Experimental source: neuroblastoma cell

C:Comment: The molecular mass is 28,000 and the pI is 5.0.

C:Keywords: brain; thiolester hydrolase

Query Match 32.9%; Score 28; DB 2; Length 15;
Best Local Similarity 46.7%; Pred. No. 89;
Matches 7; Conservative 3; Mismatches 3; Indels 2; Gaps 1;

OY 1 MNLVPM--NPLVME 13
| | | | |
Db 1 MQLKPMENPEMLNK 15

RESULT 4
50356

118k stomach cancer antigen - human (fragment)
C:Species: Homo sapiens (man)

C:Date: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 17-Mar-1999
C:Accession: A60356

R:Shiraiishi, Y.

Int. J. Cancer 45, 783-787, 1990

A:Title: Western blotting analysis for malignant lymphoma and stomach cancer antigens fr
A:Reference number: A60356; MUID:90216080

A:Accession: A60356

A:Molecule type: protein

A:Residues: 1-9 <SHI>

C:Keywords: glycoprotein

Query Match 27.1%; Score 23; DB 2; Length 9;
Best Local Similarity 28.6%; Pred. No. 2,2e+05;
Matches 2; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

OY 4 VPNNPLV 10
| | | | |
Db 1 IPLKPIVI 7

RESULT 5
S09024

carboxylesterase (EC 3.1.1.1), microsomal - golden hamster (fragment)
C:Species: Mesocricetus auratus (golden hamster)
C:Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 28-Apr-1993
C:Accession: S09024

R:Hosokawa, M.; Maki, T.; Satoh, T.

Arch. Biochem. Biophys. 277, 219-227, 1990

A:Title: Characterization of molecular species of liver microsomal carboxylesterases
A:Reference number: S09021; MUID:90179180

A:Accession: S09024

A:Molecule type: protein

A:Residues: 1-11 <HOS>

C:Keywords: carboxylic ester hydrolase

Query Match 25.9%; Score 22; DB 2; Length 11;
Best Local Similarity 50.0%; Pred. No. 6,5e+02;
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 5 PMNPLVN 12
| | | | |
Db 2 PSXPXVYN 9

RESULT 6

A48973
glucoamylase A1 (EC 3.2.1.-) - Chalara paradoxa (fragment)

C:Species: Chalara paradoxa

C:Date: 19-Dec-1993 #sequence_revision 25-Apr-1997 #text_change 25-Apr-1997
C:Accession: A48973

R:Konuma, M.; Kalnina, K.

Carbohydr. Res. 227, 385-388, 1992

A:Title: Heterogeneity of the glucoamylase components of the raw-starch-digesting amy

A:Reference number: A48973; MUID:92361881

A:Accession: A48973

A:Molecule type: protein

A:Residues: 1-11 <MON>

A:Note: sequence extracted from NCBI backbone (NCBIP:110946)

C:Keywords: glycosidase; hydrolase

Query Match 24.7%; Score 21; DB 2; Length 11;
Best Local Similarity 50.0%; Pred. No. 9,6e+02;
Matches 3; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 11 MNEEFVN 16
| | | | |
Db 3 LDEFVN 8

RESULT 7
S29859
gene p10 protein - Choristoneura fumiferana nuclear polyhedrosis virus (fragment)
C:Species: Choristoneura fumiferana nuclear polyhedrosis virus, CFNPV

C:Date: 19-Mar-1997 #sequence_revision 24-Mar-1999 #text_change 24-Mar-1999
C:Accession: S29859

R:Hill, J.E.; Kuzio, J.; Wilson, J.A.; Mackinnon, E.A.; Faulkner, P.

Biochim. Biophys. Acta 1172, 187-189, 1993

A:Title: Nucleotide sequence of the p74 gene of a baculovirus pathogenic to the spruc

A:Reference number: S29849; MUID:93176808

A:Accession: S29859

A:Molecule type: DNA

A:Status: not compared with conceptual translation

A:Residues: 1-12 <HIL>

A:Experimental source: strain Ireland

Query Match 24.7%; Score 21; DB 2; Length 12;
Best Local Similarity 80.0%; Pred. No. 1,1e+03;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 5 PMNPL 9
| | | | |

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 6, 2002, 13:23:02 ; Search time 78.06 Seconds
(without alignments)
147,968 Million cell updates/sec

Title: US-09-897-042-17

Perfect score: 51
Sequence: 1 atgaactgtgacgatgaa.....tgaacgaatcatgaactaa 51

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 498480

Minimum DB seq length: 0
Maximum DB seq length: 51

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : Issued_Patents_NA:*

- 1: /cgn2_6/prodata/2/ina/5A_COMB.seq:*
- 2: /cgn2_6/prodata/2/ina/5B_COMB.seq:*
- 3: /cgn2_6/prodata/2/ina/5A_COMB.seq:*
- 4: /cgn2_6/prodata/2/ina/5B_COMB.seq:*
- 5: /cgn2_6/prodata/2/ina/PCTUS_COMB.seq:*
- 6: /cgn2_6/prodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	18	35.3	41	4	US-09-030-156-21
2	18	35.3	48	4	US-09-358-972-150
3	16	31.4	46	5	PCT-US95-11684-24
4	16	31.4	47	2	US-08-305-764C-26
5	15.6	30.6	25	4	US-09-286-682A-9
6	15.4	29.2	36	4	US-09-255-368-38
7	14.8	29.0	27	1	US-08-758-306-868
8	14.8	29.0	28	2	US-08-666-493-18
9	14.8	29.0	28	2	US-08-666-493-23
10	14.8	29.0	31	1	US-08-485-284A-13
11	14.8	29.0	32	3	US-08-997-803-10
12	14.8	29.0	32	3	US-08-845-546-17
13	14.8	29.0	46	2	US-08-465-380-127
14	14.8	29.0	46	2	US-08-486-397-127
15	14.8	29.0	46	2	US-08-486-399-127
16	14.8	29.0	46	2	US-08-461-965-127
17	14.8	29.0	46	2	US-08-634-641-127
18	14.8	29.0	46	3	US-09-249-471-127
19	14.8	29.0	46	3	US-09-249-472-127
20	14.8	29.0	46	3	US-09-249-451-127
21	14.8	29.0	46	3	US-08-809-455-127
22	14.8	29.0	46	3	US-09-249-461-127
23	14.8	29.0	46	3	US-09-249-448-127
24	14.8	29.0	51	3	US-09-930-798-5
25	14.8	29.0	51	4	US-09-502-778A-5
26	14.6	28.6	40	1	US-08-032-846-35
27	14.6	28.6	43	1	US-07-995-950-6

28	14.6	28.6	43	1	US-08-300-582-6	Sequence 6, Appl
29	14.6	28.6	49	1	US-08-351-413-15	Sequence 15, Appl
30	14.6	28.6	49	2	US-09-025-583-15	Sequence 15, Appl
31	14.4	28.2	37	1	US-08-244-378A-19	Sequence 19, Appl
32	14.4	28.2	38	2	US-08-782-760-4	Sequence 4, Appl
33	14.4	28.2	38	5	PCT-US96-00995-4	Sequence 4, Appl
34	14.2	27.8	24	3	US-08-483-316-3	Sequence 3, Appl
35	14.2	27.8	24	5	PCT-US95-12624-3	Sequence 3, Appl
36	14.2	27.8	28	3	US-08-740-784-3	Sequence 3, Appl
37	14.2	27.8	36	2	US-08-669-721-1	Sequence 1, Appl
38	14.2	27.8	36	4	US-09-189-344-1	Sequence 1, Appl
39	14.2	27.8	41	4	US-09-143-634-13	Sequence 13, Appl
40	14.2	27.8	44	4	US-08-850-744-7	Sequence 7, Appl
41	14.2	27.8	44	4	US-08-850-744-8	Sequence 8, Appl
42	14.2	27.8	45	6	5474933-5	Patent No. 5474933
43	14.2	27.8	47	1	US-08-170-588-5	Sequence 5, Appl
44	14.2	27.8	47	2	US-08-482-182-5	Sequence 5, Appl
45	14.2	27.8	47	2	US-08-482-182-73	Sequence 73, Appl
46	14.2	27.8	50	2	US-08-821-559A-19	Sequence 19, Appl
47	14.2	27.5	30	3	US-08-777-708C-10	Sequence 10, Appl
48	14.2	27.5	36	2	US-09-018-576-10	Sequence 10, Appl
49	14.2	27.5	36	3	US-09-248-137-10	Sequence 10, Appl
50	14.2	27.5	41	4	US-09-030-156-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1
US-09-030-156-21
; Sequence 21, Application US/09030156
; Patent No. 6207373
; GENERAL INFORMATION:
; APPLICANT: Sosnowski, Ronald G.
; TITLE OF INVENTION: METHODS AND APPARATUS FOR DETERMINATION OF LENGTH
; FILE REFERENCE: DAVID B. MURPHY: Nanogen 221/157
; CURRENT APPLICATION NUMBER: US/09/030.156
; CURRENT FILING DATE: 1998-02-25
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: Patentl Ver. 2.0
; SEQ ID NO 21
; LENGTH: 41
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Human /
; OTHER INFORMATION: Bioclin
US-09-030-156-21

Query Match 35.3% Score 18; DB 4; Length 41;
Best Local Similarity 70.6% Pred. No. 1e+02;
Matches 24; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 18 gaaccactagtaatgaacgaatcatgaactaa 51
Db 6 gaaccctcactgaatgaatgaatgaatgaatga 39

RESULT 2
US-09-358-972-150
; Sequence 150, Application US/09358972
; Patent No. 6235480
; GENERAL INFORMATION:
; APPLICANT: Shultz, John W.
; APPLICANT: Lewis, Martin K.
; APPLICANT: Liepke, Donna
; APPLICANT: Mandrekar, Michelle
; APPLICANT: Kephardt, Daniel
; APPLICANT: Rhodes, Richard B.
; APPLICANT: Andrews, Christine A.
; APPLICANT: Hartnett, James R.

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 6, 2002, 14:05:27 ; Search time 143.66 Seconds
(without alignments)
304.355 Million cell updates/sec

Title: us-09-897-042-17

Perfect score: 51
Sequence: 1 atgaactgtgacccatgaa.....tgaacgaattcattgactaa 51

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1000800

Minimum DB seq length: 0
Minimum DB seq length: 51

Post-processing: Maximum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database:

N_Geneseq_1101:*

1:	/SID2/gcgdata/geneseq/geneseq/NA1980.DAT:*
2:	/SID2/gcgdata/geneseq/geneseq/NA1981.DAT:*
3:	/SID2/gcgdata/geneseq/geneseq/NA1982.DAT:*
4:	/SID2/gcgdata/geneseq/geneseq/NA1983.DAT:*
5:	/SID2/gcgdata/geneseq/geneseq/NA1984.DAT:*
6:	/SID2/gcgdata/geneseq/geneseq/NA1985.DAT:*
7:	/SID2/gcgdata/geneseq/geneseq/NA1986.DAT:*
8:	/SID2/gcgdata/geneseq/geneseq/NA1987.DAT:*
9:	/SID2/gcgdata/geneseq/geneseq/NA1988.DAT:*
10:	/SID2/gcgdata/geneseq/geneseq/NA1989.DAT:*
11:	/SID2/gcgdata/geneseq/geneseq/NA1990.DAT:*
12:	/SID2/gcgdata/geneseq/geneseq/NA1991.DAT:*
13:	/SID2/gcgdata/geneseq/geneseq/NA1992.DAT:*
14:	/SID2/gcgdata/geneseq/geneseq/NA1993.DAT:*
15:	/SID2/gcgdata/geneseq/geneseq/NA1994.DAT:*
16:	/SID2/gcgdata/geneseq/geneseq/NA1995.DAT:*
17:	/SID2/gcgdata/geneseq/geneseq/NA1996.DAT:*
18:	/SID2/gcgdata/geneseq/geneseq/NA1997.DAT:*
19:	/SID2/gcgdata/geneseq/geneseq/NA1998.DAT:*
20:	/SID2/gcgdata/geneseq/geneseq/NA1999.DAT:*
21:	/SID2/gcgdata/geneseq/geneseq/NA2000.DAT:*
22:	/SID2/gcgdata/geneseq/geneseq/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	19.6	38.4	31	15	AA065370
c 2	18.1	35.3	48	21	AA086939
c 3	17.8	34.9	49	22	AA088849
c 4	17.8	34.9	49	22	AA088850
c 5	17.6	34.5	43	18	AA049063
c 6	17.4	34.1	41	21	AA054901
c 7	16.6	31.4	47	20	AA055594
c 8	16.6	31.4	47	21	AA067749
c 9	16.6	31.4	47	21	AA067749
c 10	15.8	31.0	40	21	AA063948
c 11	15.8	31.0	41	19	AA050855

12	15.8	31.0	46	20	AA035823
13	15.8	31.0	46	21	AA089729
c 14	15.6	30.6	25	13	AA034341
c 15	15.6	30.6	25	20	AA079559
c 16	15.6	30.6	33	22	AA084503
c 17	15.6	30.6	47	21	AA069492
c 18	15.6	30.6	51	22	AA037212
c 19	15.4	30.2	27	19	AA097099
c 20	15.4	30.2	36	21	AA029469
c 21	15.4	30.2	41	19	AA047713
c 22	15.4	30.2	44	21	AA048429
c 23	15.4	30.2	49	21	AA059450
c 24	15.2	29.8	25	21	AA087733
c 25	15.2	29.8	47	21	AA067761
c 26	15.2	29.8	24	22	AA048391
c 27	15.2	29.4	30	21	AA038846
c 28	15.2	29.4	32	20	AA090651
c 29	15.2	29.4	39	18	AA006824
c 30	15.2	29.4	40	13	AA020906
c 31	14.8	29.0	27	19	AA094319
c 32	14.8	29.0	29	22	AA033907
c 33	14.8	29.0	31	14	AA048596
c 34	14.8	29.0	32	19	AA039168
c 35	14.8	29.0	32	20	AA088758
c 36	14.8	29.0	40	15	AA062481
c 37	14.8	29.0	44	20	AA021248
c 38	14.8	29.0	46	17	AA041490
c 39	14.8	29.0	46	20	AA021049
c 40	14.8	29.0	46	21	AA073630
c 41	14.8	29.0	48	21	AA061563
c 42	14.6	28.6	30	21	AA095446
c 43	14.6	28.6	37	21	AA029429
c 44	14.6	28.6	38	22	AA096226
c 45	14.6	28.6	40	15	AA073561
c 46	14.6	28.6	40	17	AA043021
c 47	14.6	28.6	40	17	AA043064
c 48	14.6	28.6	40	22	AA026047
c 49	14.6	28.6	44	18	AA070891
c 50	14.6	28.6	44	22	AA057071

ALIGNMENTS

RESULT 1
ID AA065370/c
AA065370 standard; DNA; 31 BP.

AA065370:
10-JAN-1995 (first entry)

DE Fish gonadotropic hormone beta IC chain primer.
XX
XX Polymerase chain reaction; PCR; amplify; primer; alpha; beta I;
KW beta II; chain; fish; gonadotropin hormone; GTH; oviposition;
KW animal; promotion of maturation; ss.
XX
XX Synthetic.
OS
XX
XX JP06107689-A.
XX
XX 19-APR-1994.
PD
XX
XX 20-NOV-1991; 91JP-0354152.
PF
XX
XX 20-NOV-1991; 91JP-0354152.
PR
XX
XX (NIOC) NIPPON OIL CO LTD.
PA
XX
XX WPI: 1994-163941/20.
DR
XX
XX Fish gonadotropic hormone - and transformants producing it,
PT

```

PT useful for maturation and oviposition promotion in fish
XX
PS
XX Example 5; Page 6; 22pp; Japanese.
XX
XX The sequences given in AA065365-72 are primers which were used in the
CC amplification of alpha and beta chain coding region fragments of the
CC fish gonadotropin hormone (GTH) cDNA. GTH may be produced by
CC transforming a host organism with the amplified sequence and collecting
CC the protein from the culture medium. The GTH produced may be used for
CC the promotion of maturation and oviposition in fish and other animals.
CC See also AA065373-76.
XX
XX Sequence 31 BP; 7 A; 6 C; 5 G; 13 T; 0 other:
SO

Query Match 38.4%; Score 19.6; DB 15: Length 31:
Best Local Similarity 84.6%; Pred. No. 1.3e+02:
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
XX
XX 17 tgaaccactagtaatgaacgaatc 42
XX | | | | | | | | | | | | | | | |
XX 28 TAAACCCACTGAAATGTAAGAAATTC 3
XX
XX RESULT 2
XX ID AAAB6939 standard; DNA; 48 BP.
XX AC AAAB6939;
XX XX
XX 15-JAN-2001 (first entry)
XX DE
XX TPOX allele oligonucleotide.
XX KW
XX Detection; nucleic acid hybrid; depolymerisation; analysis; SNP;
XX single nucleotide polymorphism; identification; viral load; probe;
XX genotyping; medical marker diagnostic; primer; target; mutation;
XX genetic disease; ss.
XX KW
XX Synthetic.
XX OS
XX WO200049180-A1.
XX PN
XX 24-AUG-2000.
XX PD
XX 18-FEB-2000; 2000WO-US04242.
XX XX
XX 18-FEB-1999; 99US-0252436.
XX 21-JUL-1999; 99US-0358972.
XX 25-AUG-1999; 99US-0383316.
XX PR
XX (PROM-) PROMEGA CORP.
XX PA
XX Shultz JW, Lewis MK, Lelappe D, Mandrekar M, Kephart D, Rhodes RB;
XX Andrews CA, Hartnett JR, Gu T, Olson RJ, Wood KV, Welch R;
XX WI: 2000-565377/52.
XX DR
XX
XX Determining presence or absence of a predetermined endogenous nucleic
PT acid sequence by using an enzyme that depolymerizes the 3' end of an
PT oligonucleotide probe hybridized to a target sequence to release
PT identifier nucleotides -
PT
XX
XX Example: Page 338; 389pp; English.
XX
XX The present invention describes a method (M1) for determining the
CC presence or absence of a predetermined endogenous nucleic acid target
CC sequence (ENAT). The method comprises hybridising a probe having an
CC identifier nucleotide (IN) with ENAT which is treated with an enzyme
CC that depolymerises the 3' end of hybridised NA to release the INs.
CC M1 is used for determining the number of known sequence repeats present
CC in a nucleic acid target sequence in a nucleic acid sample. The method
CC is also useful for determining whether a nucleic acid target sequence in

```

```

CC a sample is an allele from a homozygous or heterozygous locus. The
CC method is also useful for detection of mutations, translocations and
CC SNPs in nucleic acids (including those associated with genetic disease),
CC determination of viral load, species identification, sample
CC contamination, and analysis of forensic samples. AAB6791 to AAB87079
CC and AAB12617 represent sequence which are used in the exemplification of
CC the present invention.
CC N.B. there is a discrepancy between the SEQ ID NO: and sequences given
CC in the examples, and the SEQ ID NO: and sequences given in the sequence
CC listing from the present invention.
CC
XX Sequence 48 BP; 18 A; 7 C; 11 G; 12 T; 0 other;
SQ
Query Match 35.3%; Score 18; DB 21; Length 48;
Best Local Similarity 70.6%; Pred. No. 5.7e+02;
Matches 24; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
Oy 18 gaaccactagtagaacgaatcatgaactaa 51
| | | | | | | | | | | | | | | | |
Db 11 gaaccctcatgcatgaatgaatgaatgaatgaa 44
RESULT 3
AAC88849/c
ID AAC88849 standard; DNA: 49 BP.
XX
XX AAC88849;
XX
XX 05-MAR-2001 (first entry)
XX
XX Linker oligonucleotide HindXba+.
DE
XX Adenovirus type 35; Ad35; adenovirus type 5; Ad5; gene delivery vehicle;
XX gene therapy; linker; ss.
XX
XX Synthetic.
OS
XX
XX EPI054064-A1.
XX
XX 22-NOV-2000.
XX
XX 16-MAY-2000; 2000EP-0201738.
XX
XX 17-MAY-1999; 99EP-0201545.
XX
XX (INTR-) INTRIGENE BV.
XX
XX Bout A, Vogels R, Havenga MJE;
XX
XX WPI; 2001-001097/01.
XX
XX Adenovirus derived gene delivery vehicles comprising at least one
XX element of adenovirus type 35, efficiently transfers genetic material
XX to a human cell without the inherent problem of pre-existing immunity -
XX
XX Example 7; Page 20; 138BP; English.
XX
XX The present sequence is a linker used in a gene delivery vehicle
XX comprising an element of adenovirus type 35 or a functional equivalent
XX of such an element. The element is responsible for avoiding or reducing
XX neutralising activity against adenoviral elements by the host to which
XX the gene is to be delivered. The vehicle can be used to deliver genes or
XX nucleic acids of interest to host cells. Use of the delivery system
XX efficiently transfers genetic material to a human cell without the
XX inherent problem of pre-existing immunity, found with previous serotypes.
XX
XX Sequence 49 BP; 14 A; 13 C; 11 G; 11 T; 0 other;
SQ
Query Match 34.9%; Score 17.8; DB 22; Length 49;
Best Local Similarity 67.6%; Pred. No. 6.8e+02;
Matches 25; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 6, 2002, 13:18:52 ; Search time 1369.35 Seconds

(without alignments)
400.215 Million cell updates/sec

Title: us-09-897-042-17

Perfect score: 51
Sequence: 1 atgaacttggtaccgatgaa.....tgaacgaatcattgaactaa 51

Scoring table: IDENTITY-NUC
Gapop 10.0, Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 83076

Minimum DB seq length: 0
Maximum DB seq length: 51

Post-processing: Minimum Match 08
Maximum Match 100%

Listing first 50 summaries

Database :

EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estlin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_hlc:*
10: qb_est1:*
11: qb_est2:*
12: qb_hlc:*
13: qb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_tod:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	16.4	32.2	43	10	AA746752 nx63e01.s
2	15.8	31.0	28	13	AZ663964 IM0543G20
3	15.6	30.6	49	10	AJ237149 AJ237149
4	15.4	30.2	36	13	AZ427281 IM0209H10
5	15.4	30.2	43	13	AZ310736 IM0025016
6	15.2	29.8	40	13	AZ973766 2M0248C07
7	15.2	29.8	49	10	AA254733 mz83g09.r
8	15.2	29.4	43	10	AI868103 to96a07.x
9	14.8	29.0	29	13	AZ813429 2M0080011
10	14.8	29.0	32	13	AZ477577 IM0297F09
11	14.6	28.6	37	10	AA954354 ON54g01.s
12	14.6	28.6	39	13	AZ785720 2M0029G20

13	14.6	28.6	39	13	AZ827624	AZ827624	2M0104J12
14	14.6	28.6	51	13	AZ766562	AZ766562	1M0564A18
15	14.4	28.2	25	13	AZ666829	AZ666829	1M0549C17
16	14.4	28.2	45	13	AZ309922	AZ309922	1M0017C23
17	14.4	28.2	46	10	AA500769	AA500769	1M010B3.r
18	14.4	28.2	48	13	AZ377335	AZ377335	1M0131D09
19	14.4	28.2	49	10	A1327470	A1327470	ma80h03.y
20	14.4	27.8	49	11	BG151815	BG151815	na66a4h05
21	14.2	27.8	50	10	AU104069	AU104069	AU104069
22	14.2	27.5	50	10	AM643187	AM643187	cm27a03.w
23	14.2	27.5	40	10	BE217253	BE217253	258--T7 R.
24	14.2	27.5	44	11	W25663	W25663	zc64e08.r1
25	14.2	27.5	45	13	AZ772088	AZ772088	1M0574P09
26	14.2	27.5	47	13	AZ974877	AZ974877	2M0249N10
27	14.2	27.5	50	10	AU102367	AU102367	AU102367
28	13.8	27.1	27	13	AZ598057	AZ598057	1M0412B09
29	13.8	27.1	36	13	AZ650257	AZ650257	1M0520018
30	13.8	27.1	38	13	AZ800974	AZ800974	2M0059E16
31	13.8	27.1	42	13	AZ445143	AZ445143	1M0240E13
32	13.8	27.1	50	11	BG271426	BG271426	na150f04
33	13.6	26.7	24	13	AZ330762	AZ330762	1M0056L12
34	13.6	26.7	39	13	AZ640839	AZ640839	1M0503L05
35	13.6	26.7	41	13	AZ768121	AZ768121	1M0567N20
36	13.6	26.7	43	11	D20671	D20671	HUMGSO1647
37	13.6	26.7	51	11	BF507243	BF507243	5923P-1 P
38	13.6	26.7	51	13	AZ441063	AZ441063	1M0232E06
39	13.4	26.3	40	13	AZ439838	AZ439838	1M0230H07
40	13.4	26.3	44	13	AZ486826	AZ486826	1M0315I09
41	13.4	26.3	46	13	AZ498067	AZ498067	1M0335J12
42	13.4	26.3	50	10	AU106560	AU106560	AU106560
43	13.2	25.9	23	13	AZ339880	AZ339880	1M0071K01
44	13.2	25.9	23	13	AZ371086	AZ371086	1M0122D02
45	13.2	25.9	32	13	AZ793379	AZ793379	2M0046P01
46	13.2	25.9	34	10	A1583535	A1583535	ts12d01.x
47	13.2	25.9	36	13	AZ658185	AZ658185	1M0534M18
48	13.2	25.9	43	13	AZ607760	AZ607760	1M0430P22
49	13.2	25.9	49	10	AA180620	AA180620	MBAPCG1B1
50	13.2	25.9	50	11	BG152210	BG152210	na974b02

ALIGNMENTS

RESULT 1
AA746752 43 bp mRNA 27-JAN-1998
LOCUS nx63e01.s1 NCI-CGAP_Alv1 Homo sapiens CDNA clone IMAGE:1266936
DEFINITION similar to SW:ATPE_HUMAN P00846 ATP SYNTHASE A CHAIN ;, mRNA
sequence.

ACCESSION AA746752.1 GI:2786738
VERSION AA746752
KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 43)

AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

JOURNAL Tumor Gene Index

COMMENT Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: cgaps-r@mail.nih.gov

Tissue Procurement: Lee Helman, M.D., Michael R. Emmert-Buck, M.D.,

Ph.D.

CDNA Library Preparation: David B. Kitzman, Ph.D.

DNA Sequencing by: Greg Lennon, Ph.D.

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

www-bio.llnl.gov/bdrp/image/image.html

Trace considered overall poor quality

GenCore version 4.5
Copyright (c) 1993 - 2000 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 6, 2002, 13:20:22 ; Search time 1443.27 Seconds
(without alignments)
582.951 Million cell updates/sec

Title: US-09-897-042-17

Perfect score: 51

Sequence: 1 atgaactgtgaccatgaa.....tgaacgaattcatgaactaa 51

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 562506

Minimum DB seq length: 0
Maximum DB seq length: 51
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : GenEmbl:*
1: gb_da:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_da:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_om:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vl:*
30: em_hugo_hum:*
31: em_hugo_inv:*
32: em_htgo_rod:*
33: em_htg_hum:*
34: em_htg_inv:*
35: em_htg_rod:*
36: em_htg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	18	35.3	41	6	AR139408	AR139408 Sequence
2	18	35.3	48	6	AR153148	AR153148 Sequence
3	17.8	34.9	49	6	AX049948	AX049948 Sequence
4	17.8	34.9	49	6	AX049949	AX049949 Sequence
5	17.8	34.9	50	14	PGCR4G1	AF189255 Porcine g
6	17	33.3	39	6	I06895	I06895 Sequence 4
7	16.8	32.9	51	6	AX158274	AX158274 Sequence
8	16	31.4	47	6	AR026322	AR026322 Sequence
9	15.8	31.0	46	6	E32392	E32392 Thermostabl
10	15.6	30.6	33	6	AX113720	AX113720 Sequence
11	15.6	30.6	51	6	AX114885	AX114885 Sequence
12	15.4	30.2	51	6	AX157029	AX157029 Sequence
13	15.4	30.2	51	6	AX157031	AX157031 Sequence
14	15.4	30.2	51	6	AX157032	AX157032 Sequence
15	15.2	29.8	47	6	AX114386	AX114386 Sequence
16	15.2	29.8	51	6	AX158273	AX158273 Sequence
17	15.2	29.8	51	10	MUSBMP24M	D45014 Mouse gene
18	15	29.4	39	6	A66657	A66657 Sequence 2
19	14.8	29.0	27	6	AR040020	AR040020 Sequence
20	14.8	29.0	28	6	A50939	A50939 Sequence 18
21	14.8	29.0	28	6	A50944	A50944 Sequence 23
22	14.8	29.0	28	6	AR072776	AR072776 Sequence
23	14.8	29.0	28	6	AR072781	AR072781 Sequence
24	14.8	29.0	31	6	AR007177	AR007177 Sequence
25	14.8	29.0	32	6	AR099746	AR099746 Sequence
26	14.8	29.0	40	6	E07012	E07012 Primer 9/1
27	14.8	29.0	46	6	AR031892	AR031892 Sequence
28	14.8	29.0	46	6	AR031948	AR031948 Sequence
29	14.8	29.0	46	6	AR036329	AR036329 Sequence
30	14.8	29.0	46	6	AR075014	AR075014 Sequence
31	14.8	29.0	46	6	AR103878	AR103878 Sequence
32	14.8	29.0	51	6	AR119105	AR119105 Sequence
33	14.6	28.6	34	9	S80814	S80814 gamma delta
34	14.6	28.6	40	6	AX069262	AX069262 Sequence
35	14.6	28.6	43	6	I63348	I63348 Synthetic o
36	14.6	28.6	43	6	A21364	A21364 Synthetic o
37	14.6	28.6	43	6	AR007507	AR007507 Sequence
38	14.6	28.6	44	6	AX081683	AX081683 Sequence
39	14.6	28.6	49	6	AR007525	AR007525 Sequence
40	14.6	28.6	49	6	AR084091	AR084091 Sequence
41	14.6	28.6	51	6	AX165586	AX165586 Sequence
42	14.4	28.2	37	6	AR009554	AR009554 Sequence
43	14.4	28.2	38	6	AR072910	AR072910 Sequence
44	14.4	28.2	48	6	AX062298	AX062298 Sequence
45	14.4	28.2	50	10	AF071704	AF071704 Mus muscu
46	14.4	28.2	51	6	AX115073	AX115073 Sequence
47	14.4	28.2	51	6	AX117017	AX117017 Sequence
48	14.2	27.8	28	6	AR097982	AR097982 Sequence
49	14.2	27.8	36	6	AR053365	AR053365 Sequence
50	14.2	27.8	36	6	AR131089	AR131089 Sequence

ALIGNMENTS

RESULT 1
LOCUS AR139408 41 bp DNA
DEFINITION Sequence 21 from patent US 6207373.
ACCESSION AR139408
VERSION AR139408.1 GI:14481904
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 41)
AUTHORS Sosnowski, R.G. and Tu, E.
TITLE Methods for determining nature of repeat units in DNA
JOURNAL Patent: US 6207373-A 21 27-MAR-2001.
FEATURES Location/Qualifiers

RESULT 2
 5427925-27
 Patent No. 5427925
 APPLICANT: GEARING, DAVID P.; GOUGH, NICHOLAS M.; HILTON,
 DOUGLAS J.; KING, JULIE A.; METCALE, DONALD; NICE, EDOUARD C.;
 NICOLA, NICOS A.; STIMPSON, RICHARD J.; WILLSON, TRACY A.
 TITLE OF INVENTION: RECOMBINANT METHOD FOR MAKING
 LEUKEMIA INHIBITOR FACTOR
 NUMBER OF SEQUENCES: 38
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/58,979
 FILING DATE: 06-MAY-1993
 APPLICATION NUMBER: 948,614
 FILING DATE: 22-SEP-1992
 APPLICATION NUMBER: 667,159
 FILING DATE: 11-MAR-1991
 SEQ ID NO: 27
 LENGTH: 27
 5427925-27

Query Match 16.4%; Score 43; DB 6; Length 27;
Best Local Similarity 29.6%; Pred. No. 28;
Matches 8; Conservative 6; Mismatches 13; Indels 0; Gaps 0;

OY 2 NLVPRNPLVMNGFCRYPSHMRPLEDIR 28
Db 1 NPLPTPYXATXAIRHPCGNLMNQIR 27

RESULT 3

US-08-310-912A-66
; Sequence 66, Application US/08310912A
; Patent No. 5981730
; GENERAL INFORMATION:
; APPLICANT: Ausubel, Frederick M.
; APPLICANT: Staskawicz, Brian J.
; APPLICANT: Brent, Andrew F.
; APPLICANT: Dahlbeck, Douglas
; APPLICANT: Katagiri, Fumiaki
; APPLICANT: Kunkel, Barbara N.
; APPLICANT: Mindrinos, Michael N.
; APPLICANT: Yu, Guo-Liang
; TITLE OF INVENTION: RPS2 GENE FAMILY, PRIMERS, PROBES, AND DETECTION
; TITLE OF INVENTION: METHODS
; NUMBER OF SEQUENCES: 208
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2904
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/310.912A
; FILING DATE: September 22, 1994
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/227,360
; FILING DATE: April 13, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Lech, Karen F.
; REGISTRATION NUMBER: 35,238
; REFERENCE/DOCKET NUMBER: 00786/254001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 100254
; INFORMATION FOR SEQ ID NO: 66:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 35 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-310-912A-66

Query Match 16.0%; Score 42; DB 2; Length 35;
Best Local Similarity 50.0%; Pred. No. 50;
Matches 6; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 10 VMNGFCRYPSHW 21
Db 24 VOEGFCHIEHW 35

RESULT 4

US-08-841-089-66
; Sequence 66, Application US/08841089
; Patent No. 6127607
; GENERAL INFORMATION:
; APPLICANT: Ausubel, Frederick M.
; APPLICANT: Staskawicz, Brian J.
; APPLICANT: Brent, Andrew F.
; APPLICANT: Dahlbeck, Douglas
; APPLICANT: Katagiri, Fumiaki
; APPLICANT: Kunkel, Barbara N.
; APPLICANT: Mindrinos, Michael N.
; APPLICANT: Yu, Guo-Liang
; TITLE OF INVENTION: RPS2 GENE AND USES THEREOF
; NUMBER OF SEQUENCES: 106
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street Suite 3100
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2904
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/841.089
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/227,360
; FILING DATE: 13-APR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 00786/230001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 100254
; INFORMATION FOR SEQ ID NO: 66:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 35 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-841-089-66

Query Match 16.0%; Score 42; DB 3; Length 35;
Best Local Similarity 50.0%; Pred. No. 50;
Matches 6; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 10 VMNGFCRYPSHW 21
Db 24 VOEGFCHIEHW 35

RESULT 5

US-09-301-085-66

; Sequence 66, Application US/09301085
; Patent No. 6262248
; GENERAL INFORMATION:
; APPLICANT: Ausubel, Frederick M.
; APPLICANT: Staskawicz, Brian J.
; APPLICANT: Brent, Andrew F.
; APPLICANT: Dahlbeck, Douglas
; APPLICANT: Katagiri, Fumiaki
; APPLICANT: Kunkel, Barbara N.
; APPLICANT: Mindrinos, Michael N.
; APPLICANT: Yu, Guo-Liang

APPLICANT: Ausubel, Frederick M.
APPLICANT: Staskawicz, Brian J.
APPLICANT: Brent, Andrew F.
APPLICANT: Dahlbeck, Douglas
APPLICANT: Katagiri, Fumiaki
APPLICANT: Kunkel, Barbara N.
APPLICANT: Mindrinos, Michael N.
APPLICANT: Yu, Guo-Liang


```

PR 07-NOV-1997; 97US-0064908.
PR 07-NOV-1997; 97US-0064911.
PR 07-NOV-1997; 97US-0064912.
PR 07-NOV-1997; 97US-0064983.
PR 07-NOV-1997; 97US-0064984.
PR 07-NOV-1997; 97US-0064985.
PR 07-NOV-1997; 97US-0064987.
PR 07-NOV-1997; 97US-0064988.
PR 17-NOV-1997; 97US-0066094.
PR 17-NOV-1997; 97US-0066094.
PR 17-NOV-1997; 97US-0066094.
PR 17-NOV-1997; 97US-0066095.
PR 17-NOV-1997; 97US-0066089.
XX XX

(PHUA-) HUMAN GENOME SCI INC.

PI Carter KC, Ebner R, Endress GA, Feng P, Janat F;
PI Kyaw H, Lafleur DW, Moore PA, Ni J, Olsen HS, Rosen CA;
PI Ruben SM, Shi Y, Soppet DR, Wei Y;
XX N-PSDB; AAX85049.
XX MPI: 1999-337740/28.
XX N-PSDB; AAX85049.
XX

New human secreted proteins and coding sequences useful for treating
disorders of the immune system and hyperproliferative disorders

Disclosure: Page 168; 507pp; English.

XX PS This sequence represents a secreted human protein encoded by the gene
XX CC clone detailed in the descriptor line. The gene can be used to generate
XX CC fusion proteins by linking to the gene to a human immunoglobulin Fc
XX CC portion (e.g., AAX84924) for increasing the stability of the fused
XX CC protein as compared to the human protein only.
XX CC The invention relates to 125 novel genes and their fragments (nucleic
XX CC acid sequences: AAX84933-X85057; amino acid sequences AAY27567-Y27933)
XX CC which are useful for preventing, treating or ameliorating medical
XX CC conditions e.g., by protein or gene therapy. Also, pathological
XX CC conditions can be diagnosed by determining the amount of the new
XX CC polypeptides in a sample or by determining the presence of mutations in
XX CC the new polynucleotides. Specific uses are described for each of the 125
XX CC polynucleotides, based on which tissues they are most highly expressed in
XX CC (see AAX84933 for described uses).
XX XX

Sequence 28 AA:

SO Query Match 19.8%; Score 52; DB 20; Length 28;
Best Local Similarity 38.5%; Pred. No. 6.4;
Matches 10; Conservative 3; Mismatches 3; Indels 10; Gaps 2
OY 8 PLVMNGFCH---YPS-----HWRP 23
I : ||||| : |||||
Db 1 pIplngfcesreffpdsgsvllhwrp 26

RESULT 2
AAM13623
ID AAM13623 standard; Protein: 44 AA.
XX AAM13623:
XX AC
XX DT 12-OCT-2001 (first entry)
XX DE Peptide #57 encoded by probe for measuring cervical gene expression.
XX KW Probe: human: microarray; gene expression; cervical epithelial cell;
XX cervical cancer.
XX OS Homo sapiens.
XX PN W0200157278-A2.
XX PD 09-AUG-2001.
XX XX
```

```

PE      30-JAN-2001; 2001MO-US00670.
PR      04-FEB-2000; 2000US-0180312.
PR      26-MAY-2000; 2000US-0207456.
PR      30-JUN-2000; 2000US-0608408.
PR      03-AUG-2000; 2000US-0632366.
PR      21-SEP-2000; 2000US-0234687.
PR      27-SEP-2000; 2000US-0236359.
PR      04-OCT-2000; 2000GB-0024263.
XX
PA      (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI      Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR      WPI; 2001-488901/53.
XX
PT      Human genome-derived single exon nucleic acid probes useful for
XX      analyzing gene expression in human cervical epithelial cells -
XX
PS      Claim 27; SEQ ID No 18449; 487bp; English.
XX
CC      The present invention relates to human single exon nucleic acid probes
CC      (SENP: see AA110068-AA128459). The present sequence is a peptide encoded
CC      by one such probe. The SENPs are derived from human HeLa cells. The SENPs
CC      can be used to produce a single exon microarray, which can be used for
CC      measuring human gene expression in a sample derived from human cervical
CC      epithelial cells. By measuring gene expression, the probes are therefore
CC      useful in grading and/or staging of diseases of the cervix, notably
CC      cervical cancer.
CC      Note: The sequence data for this patent did not form part of the printed
CC      specification, but was obtained in electronic format directly from WIPO
CC      at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ      Sequence    44 AA:
XX
Query Match          16 6%; Score 43.5; DB 22; Length 44;
Best Local Similarity 29.4%; Pred. No. 1.3e+02;
Matches   10; Conservative   3; Mismatches 10; Indels   11; Gaps   1.
OY      18 PSHW-----RPLEQIRLTKPERRLSML 40
        | | | | | | | | | | | | | | | |
DB      5 pgrwrhcarnrtgrrcrpwwqslsvprpsrpgwl 38
XX
RESULT      3
AAM26020
ID      AAM26020 standard; Protein: 44 AA.
XX
AC      AAM26020;
XX
DT      17-OCT-2001 (first entry)
XX
DE      Peptide #57 encoded by probe for measuring placental gene expression.
XX
KW      Probe; microarray; human; placenta; antenatal diagnosis;
XX      genetic disorder.
XX
OS      Homo sapiens.
XX
PN      WO200157272-A2.
XX
PD      09-AUG-2001.
PF      30-JAN-2001; 2001WO-US00663.
XX
PR      04-FEB-2000; 2000US-0180312.
PR      26-MAY-2000; 2000US-0207456.
PR      30-JUN-2000; 2000US-0608408.
PR      03-AUG-2000; 2000US-0632366.
PR      21-SEP-2000; 2000US-0234687.
PR      27-SEP-2000; 2000US-0236359.
PR      04-OCT-2000; 2000GB-0024263.

```


RA Jatali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laok P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Maitel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Mekulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacled J.M.,
RA Palazzolo M., Piltman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA She B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spler E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svrskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhou W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster";
Science 287:2185-2195(2000).
EMBL: AE003799; AAE57708.1; -.
FlyBase: FBgn0025583; IM2.
DR
SQ SEQUENCE 45 AA: 4740 MW: 81986CD6388135B7 CRC64;

Query Match 17.9% Score 47; DB 5; Length 45;
Best Local Similarity 45.0% Pred. No. 36;
Matches 9; Conservative 5; Mismatches 2; Indels 4; Gaps 1;
Oy 2 NLVPMNP---LVNMGFCRY 17
Db 19 NAVEPLSPGCVNVIWGCKY 38

RESULT 2
O9PV48 PRELIMINARY; PRT: 38 AA.
ID O9PV48
AC O9PV48;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE DYSTROPHIN (FRAGMENT).
GN DYST.
OS Hippocampus sp.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Gasterosteiformes;
OC Syngnathidae; Hippocampus.
NCBI_TaxID=72047;
[1]
RN
RP SEQUENCE FROM N.A.
RX MEDLINE=93398697; PubMed=10468597;
RA Venkatesh B., Ning Y., Brenner S.;
RT "Late changes in spliceosomal introns define clades in vertebrate
evolution";
Proc. Natl. Acad. Sci. U.S.A. 96:10267-10271(1999).
RL EMBL: AF137147; AAD53424.1; -.
DR EMBL: AF137147; AAD53424.1; -.
FT NON_TER 1
FT NON_TER 38
SQ SEQUENCE 38 AA: 4345 MW: AADB3DA1F62C0193 CRC64;

Query Match 17.0% Score 44.5; DB 13; Length 38;
Best Local Similarity 42.3% Pred. No. 65;
Matches 11; Conservative 2; Mismatches 4; Indels 9; Gaps 1;
Oy 31 TKPERRLSWL-----PPLSN 47
Db 3 TALEEVLTLWLSAEDGLOGPPISNN 28

RESULT 3
O9PV42 PRELIMINARY; PRT: 38 AA.
ID O9PV42

AC O9PV42;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE DYSTROPHIN (FRAGMENT).
GN DYST.
OS Siganus sp.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes;
OC Acanthuroidei; Siganidae; Siganus.
NCBI_TaxID=94227;
[1]
RN
RP SEQUENCE FROM N.A.
RX MEDLINE=93398697; PubMed=10468597;
RA Venkatesh B., Ning Y., Brenner S.;
RT "Late changes in spliceosomal introns define clades in vertebrate
evolution";
Proc. Natl. Acad. Sci. U.S.A. 96:10267-10271(1999).
RL EMBL: AF137153; AAD53430.1; -.
DR EMBL: AF137153; AAD53430.1; -.
FT NON_TER 1
FT NON_TER 38
SQ SEQUENCE 38 AA: 4315 MW: AAC1E781F62C0193 CRC64;

Query Match 17.0% Score 44.5; DB 13; Length 38;
Best Local Similarity 42.3% Pred. No. 65;
Matches 11; Conservative 2; Mismatches 4; Indels 9; Gaps 1;
Oy 31 TKPERRLSWL-----PPLSN 47
Db 3 TALEEVLTLWLSAEDGLOGPPISNN 28

RESULT 4
O9PV35 PRELIMINARY; PRT: 38 AA.
ID O9PV35
AC O9PV35;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE DYSTROPHIN (FRAGMENT).
GN DYST.
OS Ostracion sp.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Ostraciidae; Ostracion.
NCBI_TaxID=94239;
[1]
RN
RP SEQUENCE FROM N.A.
RX MEDLINE=93398697; PubMed=10468597;
RA Venkatesh B., Ning Y., Brenner S.;
RT "Late changes in spliceosomal introns define clades in vertebrate
evolution";
Proc. Natl. Acad. Sci. U.S.A. 96:10267-10271(1999).
RL EMBL: AF137160; AAD53437.1; -.
DR EMBL: AF137160; AAD53437.1; -.
FT NON_TER 1
FT NON_TER 38
SQ SEQUENCE 38 AA: 4315 MW: AAC1E781F62C0193 CRC64;

Query Match 17.0% Score 44.5; DB 13; Length 38;
Best Local Similarity 42.3% Pred. No. 65;
Matches 11; Conservative 2; Mismatches 4; Indels 9; Gaps 1;
Oy 31 TKPERRLSWL-----PPLSN 47
Db 3 TALEEVLTLWLSAEDGLOGPPISNN 28

RESULT 5
O9PV46

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OW protein - protein search, using sw model

Run on: January 6, 2002, 09:45:27 : Search time 9.85 Seconds
(without alignments)
174.949 Million cell updates/sec

Title: US-09-897-042-16

Perfect score: 262
Sequence: 1 MNLVPMNPLVNMNFCRPSH.....RLTKPERRLSWLLPLLSNN 47

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 segs, 36664827 residues

Total number of hits satisfying chosen parameters: 3287

Minimum DB seq length: 0
Maximum DB seq length: 47

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	36	13.7	32	1	CRP_PLEPL
2	35	13.4	24	1	CCAA_STRTI
3	35	13.4	41	1	CSP2_STOR
4	34	13.0	43	1	BAGE_HUMAN
5	34	13.0	43	1	PCBN_ZAMFU
6	34	13.0	46	1	YPCA_ECOLI
7	33	12.6	23	1	TL17_SPTOL
8	33	12.6	33	1	LPRH_ECOLI
9	33	12.6	36	1	PSAI_CVACA
10	33	12.6	40	1	YSXC_SULAC
11	32	12.2	42	1	NGF_VIPLE
12	31.5	12.0	38	1	BD08_BOVIN
13	31.5	12.0	40	1	BD07_BOVIN
14	31	11.8	43	1	PSBN_ARATH
15	31	11.8	43	1	PSBN_ORYSA
16	31	11.8	43	1	PSBN_PEA
17	30	11.5	25	1	ANT3_MESAU
18	30	11.5	30	1	ITR2_ECELU
19	30	11.5	36	1	PAHO_ALIMI
20	30	11.5	36	1	PAHO_ARSAN
21	30	11.5	36	1	PAHO_ERIEU
22	30	11.5	43	1	PSBN_PINTH
23	30	11.5	44	1	RS7_SALSA
24	29.5	11.3	45	1	YA78_ARCFU
25	29	11.1	34	1	AMP2_CHICK
26	29	11.1	36	1	PAHO_EOUZE
27	29	11.1	39	1	GIL2_CHICK
28	29	11.1	47	1	HSP1_CAVPO
29	28.5	10.9	37	1	SCKA_TITSE
30	28.5	10.9	39	1	SCKA_CENNA
31	28.5	10.9	43	1	CPC6_CANPG
32	28.5	10.9	43	1	YSK_CYRV
33	28.5	10.9	46	1	HSP1_ORCOR

34	28.5	10.9	47	1	H4Y_BLEJA	P80738 blepharisma
35	28	10.7	38	1	VG8_SPV4	P11340 spilioplasma
36	28	10.7	41	1	CSP_STOR	O33669 streptococc
37	28	10.7	42	1	DLPI_ORMAN	P82172 ornithorhyn
38	28	10.7	47	1	RK14_VIGUN	P26820 vigna ungu
39	27.5	10.5	35	1	IAAC_HORVU	P34951 hordeum vul
40	27.5	10.5	37	1	GHR3_RAT	P33560 rattus norv
41	27	10.3	15	1	UBL1_MONDO	P50103 monodelphis
42	27	10.3	28	1	ITR3_LORCY	P35628 iufta cylin
43	27	10.3	28	1	TXO2_AGEAP	P15971 agelenopsis
44	27	10.3	32	1	RS20_THETH	P80360 cheurus aqu
45	27	10.3	35	1	YOB5_CAEEL	O09258 caenorhabdi
46	27	10.3	36	1	TX1B_AGEAP	P15970 agelenopsis
47	27	10.3	37	1	RL36_AOUAE	O66467 aquilex aeo
48	27	10.3	37	1	VG65_BPPH2	P16515 bacteriopho
49	27	10.3	37	1	YLZ8_CAEEL	P34421 caenorhabdi
50	27	10.3	39	1	DECO_MACDE	P17350 macrodelia

ALIGNMENTS

RESULT 1				
ID	CRP_PLEPL	STANDARD:	PRT:	32 AA.
AC	P12245;			
DT	01-OCT-1989 (Rel. 12, Created)			
DT	01-OCT-1989 (Rel. 12, Last sequence update)			
DT	01-OCT-1994 (Rel. 30, Last annotation update)			
DE	C-REACTIVE PROTEIN (FRAGMENT).			
OS	Pleuronectes platessa (Plaice).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;			
OC	Acanthomorpha; Acanthopterygii; Percomorpha; Pleuronectiformes;			
OC	Pleuronectidae; Pleuronectidae; Pleuronectes.			
OX	NCBI_TaxID=8262;			
RM	[1]			
RP	SEQUENCE.			
RX	MEDLINE=82232106; PubMed=7093286;			
RA	Peyers M.B., de Beer F.C., Milstein C.P., March J.F., Feinstein A.,			
RA	Butress N., Clamp J.R., Taylor J., Bruton C., Fletcher T.C.;			
RT	"C-reactive protein and serum amyloid P component in the plaice			
RT	(Pleuronectes platessa L.), a marine teleost, are homologous with			
RT	their human counterparts."			
RL	Biochim. Biophys. Acta 704:123-133(1982).			
CC	-1- FUNCTION: CRP DISPLAYS SEVERAL FUNCTIONS ASSOCIATED WITH HOST			
CC	DEFENSE: IT PROMOTES AGGLUTINATION, BACTERIAL CAPSULAR SWEETING,			
CC	PHAGOCYTOSIS, & COMPLEMENT FIXATION THROUGH ITS CALCIUM-DEPENDENT			
CC	BINDING TO PHOSPHOTICHOLINE.			
CC	-1- SUBUNIT: HOMOPENTAMER. PENTAXIN (OR PENTAXIN) HAVE A DISCOID			
CC	ARRANGEMENT OF 5 NONCOVALENTLY BOUND SUBUNITS.			
CC	-1- SIMILARITY: BELONGS TO THE PENTAXIN FAMILY.			
DR	InterPro: IPR001759; Pentaxin.			
DR	PROSITE: PS00289; PENTAXIN; PARTIAL.			
KW	Acute phase; Calcium; Pentaxin.			
FT	DOMAIN 1			
FT	NON TER 32			
FT	SEQUENCE 32 AA; 3658 MW; EDB2E022104FA227 CRC64;			
Query Match				
Best Local Similarity		13.7%;	Score 36;	DB 1; Length 32;
Matches 6; Conservative		42.9%;	Pred. No. 2.2e+02;	
		3;	Mismatches 5;	Indels 0; Gaps 0;
OY	1 MNLVPMNPLVNMNCF 14			
DB	17 VELIPMKPLINLRAF 30			
RESULT 2				
ID	CCAA_STRTI	STANDARD:	PRT:	24 AA.

AC P80436;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE CHINOXALIN-2-CARBOXYLIC ACID ACTIVATING ENZYME (FRAGMENT).
OS Streptomyces triostinicus.
CC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
CC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
RN NCBI_TaxID=45399;
RX [1]
RP SEQUENCE.
RA Pohl A., Schlumbohm W., Keller U.;
RL Submitted (MAR-1993) to the SWISS-PROT data bank.
KW -1- FUNCTION: INVOLVED IN TRIOSTIN BIOSYNTHESIS.
CC Antibiotic biosynthesis.
FT NON TER 24 24
SQ SEQUENCE 24 AA: 2900 MW: 91C222B657CEB6D1 CRC64;

Query Match 13.4%; Score 35; DB 1; Length 24;
Best Local Similarity 29.0%; Pred. No. 2.2e+02;
Matches 9; Conservative 5; Mismatches 9; Indels 8; Gaps 1;

Qy 10 VMNGCRFPRSHWRPLEQIRLLTKPRRLSWL 40
Db 1 MLDFVPMWDH-----LADEYRRRGICWL 23

RESULT 3
CSP2_STROR STANDARD; PRT; 41 AA.
ID CSP2_STROR
AC O3690;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE COMPETENCE STIMULATING PEPTIDE PRECURSOR (CSP).
GN COMC.
OS Streptococcus oralis.
CC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
CC Streptococcus.
OX NCBI_TaxID=1303;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=DSM 20066;
RL MEDLINE=98012953; PubMed=9352904;
RX Haeverstein L.S., Hakenbeck R., Gaustad P.;
DT "Natural competence in the genus Streptococcus: evidence that streptococci can change phenotype by interspecies recombinational exchanges";
CC J. Bacteriol. 179:6589-6594(1997).
RL J. Bacteriol. 179:6589-6594(1997).
CC -1- FUNCTION: ACTS AS A PHEROMONE, INDUCES CELLS TO DEVELOP COMPETENCE FOR GENETIC TRANSFORMATION.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- SIMILARITY: BELONGS TO THE COMC FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: AJ000874; CAA04364.1; -.
KW Pheromone; Competence.
FT PROPEP 1 24
FT CHAIN 25 41
SQ SEQUENCE 41 AA: 4961 MW: B202A064BCD451AD CRC64;

Query Match 13.4%; Score 35; DB 1; Length 41;
Best Local Similarity 50.0%; Pred. No. 3.7e+02;
Matches 8; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

Qy 21 WRPLEQIRLLTKPRRR 36
Db 26 WRISSETIRLLFPRRK 41

RESULT 4
BAGE_HUMAN STANDARD; PRT; 43 AA.
ID BAGE_HUMAN
AC Q13072;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE B MELANOMA ANTIGEN (ANTIGEN M22-BA).
GN BAGE.
OS Homo sapiens (Human).
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Melanoma;
RX MEDLINE=95202592; PubMed=7895173;
RA Boel P., Wildmann C., Sensi M.L., Brasseur R., Renaud J.-C.,
RA Coulle P., Boon T., van der Bruggen P.;
RT "BAGE: a new gene encoding an antigen recognized on human melanomas by cytolytic T lymphocytes";
RL Immunity 2:167-175(1995).
CC -1- FUNCTION: NOT KNOWN. ANTIGEN RECOGNIZED ON A MELANOMA BY AUTOLOGOUS CYTOLYTIC T LYMPHOCYTES.
CC -1- TISSUE SPECIFICITY: NOT EXPRESSED IN NORMAL TISSUES, EXCEPT IN TESTIS. EXPRESSED WITH SIGNIFICANT PROPORTION IN MELANOMAS, BUT ALSO IN TUMORS OF VARIOUS HISTOLOGICAL ORIGINS, SUCH AS BLADDER CARCINOMAS, HEAD AND NECK SQUAMOUS CELL CARCINOMAS, LUNG AND BREAST CARCINOMAS, NOT EXPRESSED IN RENAL, COLORECTAL AND PROSTATIC CARCINOMAS, LEUKEMIAS AND LYMPHOMAS. MORE FREQUENTLY EXPRESSED IN METASTATIC MELANOMAS THAN IN PRIMARY MELANOMAS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: U19180; AAC50123.1; -.
DR MIM: 605167; -.
KW Antigen.
SQ SEQUENCE 43 AA: 4810 MW: 36F3BCE4012F1BB CRC64;

Query Match 13.0%; Score 34; DB 1; Length 43;
Best Local Similarity 40.0%; Pred. No. 5.2e+02;
Matches 8; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

Qy 24 LEQIRLLTKPRRLSWLPP 43
Db 14 LLQARLMKEEPVYSWRLEP 33

RESULT 5
PSBN_ZAMFU STANDARD; PRT; 43 AA.
ID PSBN_ZAMFU
AC O9MSRL;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PHOTOSYSTEM II REACTION CENTER N PROTEIN.
GN PSBN.
OS Zamia furfuracea.
CC Chloroplast.
CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 6, 2002, 09:43:22 ; Search time 18.55 Seconds

(without alignments)
193.003 Million cell updates/sec

Title: US-09-897-042-16

Perfect score: 262

Sequence: 1 MNLVPMNPLVMNGFCRYPH.....RLTKPERRLSMLPLPLSNN 47

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 10378

Minimum DB seq length: 0

Maximum DB seq length: 47

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	42	16.0	44	2	S29785
2	41	15.6	39	2	S77904
3	40	15.3	29	2	S54340
4	39	14.9	31	2	A59321
5	38	14.5	32	2	A53088
6	36	13.7	32	2	A05291
7	36	13.7	43	2	E64740
8	35	13.4	40	2	B96843
9	35	13.4	41	2	A49964
10	34.5	13.2	35	2	A61582
11	34.5	13.2	40	2	S19539
12	34	13.0	26	2	A60044
13	34	13.0	33	2	T39160
14	34	13.0	40	2	T64846
15	34	13.0	40	2	S17574
16	34	13.0	43	2	T00159
17	34	13.0	46	2	S12084
18	33.5	12.8	45	2	PC1218
19	33	12.6	30	2	A25497
20	33	12.6	33	2	A05111
21	33	12.6	33	2	D86064
22	33	12.6	40	2	S21044
23	33	12.6	43	2	I39690
24	33	12.6	46	2	A85605
25	33	12.6	46	2	B85656
26	32.5	12.4	38	2	E82858
27	32	12.2	36	2	H32502
28	32	12.2	38	2	H39690
29	32	12.2	45	2	T01784

30	32	12.2	46	2	A35437	nicotinic acetylch
31	31.5	12.0	32	2	C56649	streptomycin B-11k
32	31.5	12.0	36	2	H70251	hypothetical prote
33	31.5	12.0	40	2	G45495	beta-defensin-7 -
34	31.5	12.0	40	2	I45495	beta-defensin-9 -
35	31.5	12.0	43	2	T46409	hypothetical prote
36	31.5	12.0	45	2	H82526	hypothetical prote
37	31.5	12.0	46	2	D81077	hypothetical prote
38	31	11.8	17	2	PS0454	38k protein 3129 -
39	31	11.8	25	2	S52124	alpha-gliadin S11
40	31	11.8	26	2	D82344	IS1004 transposase
41	31	11.8	27	2	E44629	homeotic protein H
42	31	11.8	27	2	C44621	homeotic protein H
43	31	11.8	30	2	B53088	factor IX/factor X
44	31	11.8	34	2	I38214	protein-serine/thr
45	31	11.8	40	2	A34122	lectin-related pro
46	31	11.8	42	2	T07557	hypothetical prote
47	31	11.8	43	1	F2R2N	photosystem II pro
48	31	11.8	43	2	S05216	photosystem II pro
49	31	11.8	43	2	S14142	photosystem II pro
50	31	11.8	43	2	S12133	photosystem II pro

ALIGNMENTS

RESULT 1
S29785
NADH oxidase (H2O2-forming) (EC 1.6.-.-) - Thermus aquaticus
C:Species: Thermus aquaticus
C:Date: 13-Jan-1995 #sequence.revision 13-Jan-1995 #text.change 17-Mar-1999
C:Accession: S29785
R:Park, H.T.; Kreutzer, R.; Reiser, C.O.A.; Sprinzl, M.
Eur. J. Biochem. 211, 909, 1993
A:Title: Molecular cloning and nucleotide sequence of the gene encoding a H(2)O(2)-to
A:Reference number: S29785; MUID:93170329
A:Accession: S29785
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-44 <PAR>
A:Cross-references: EMBL:X60110
A:Note: the source is designated as Thermus thermophilus
C:Keywords: Nad; oxidoreductase

Query Match 16.0% Score 42; DB 2; Length 44;
Best Local Similarity 61.5%; Pred. No. 90;
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
OY 17 YPSHMRPLEQIRL 29
Db 30 YPSHRLPLERVL 42

RESULT 2
S77904
tax protein - simian immunodeficiency virus SIVagm (fragment)
C:Species: simian immunodeficiency virus SIVagm
C:Date: 19-Mar-1997 #sequence.revision 18-Jul-1997 #text.change 20-Sep-1999
C:Accession: S77904
R:Vandamme, A.M.
submitted to the EMBL Data Library, May 1994
A:Reference number: S44288
A:Accession: S77904
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-39 <VAN>
A:Cross-references: EMBL:Z32851; NID:9483445; PIDN:CAA83684.1; PID:9483446
A:Experimental source: cell-line pp 1664
A:Note: the source is designated as primate T-cell lymphotropic virus
C:Superfamily: leukemia virus trans-activating transcription regulator

Query Match 15.6%; Score 41; DB 2; Length 39;
Best Local Similarity 36.7%; Pred. No. 1.1e+02;
Matches 11; Conservative 2; Mismatches 13; Indels 4; Gaps 1;
OY 10 VVNGCFRPSHMRPLEQIRLLTKPERRLSW 39
DB 12 VSGGLCTRLRLHAL----LATCPHQLTW 37

RESULT 3
S54340
diazepam binding inhibitor 2 - human
C:Species: Homo sapiens (man)
C:Date: 27-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 07-May-1999
C:Accession: S54340
R:Kolmer, M.; Rovio, A.; Alho, H.
Biochem. J. 306, 327-330, 1995
A:Title: The characterization of two diazepam binding inhibitor (DBI) transcripts in hum
Accession: S54340; MUID:95194304
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-29 <KOL>

Query Match 15.3%; Score 40; DB 2; Length 29;
Best Local Similarity 77.8%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 39 WLPLPLSN 47
DB 6 WLPPASAN 14

RESULT 4
A59321
protein bmk 33-1 [imported] - Manchurian scorpion
C:Species: Mesobuthus martensii (Manchurian scorpion)
C:Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 02-Sep-2000
C:Accession: A59321
R:Xu, C.
submitted to the Protein Sequence Database, August 2000
A:Reference number: A59321
A:Accession: A59321
A:Status: preliminary
A:Molecule type: protein
A:Residues: 1-31 <XUC>

Query Match 14.9%; Score 39; DB 2; Length 31;
Best Local Similarity 63.6%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 5 PMNPLVMNGFC 15
DB 18 PGNPPCRNGFC 28

RESULT 5
A53088
factor IX/factor X-binding anticoagulant protein B chain - jararaca (fragment)
C:Species: Bothrops jararaca (jararaca)
C:Date: 13-Sep-1994 #sequence_revision 18-Nov-1994 #text_change 15-Jun-1996
C:Accession: A53088
R:Sekiya, F.; Atoda, H.; Morita, T.
Biochemistry 32, 6892-6897, 1993
A:Title: Isolation and characterization of an anticoagulant protein homologous to botroc
A:Reference number: A53088; MUID:93326575
A:Accession: A53088
A:Status: preliminary
A:Molecule type: protein
A:Residues: 1-30 <SEK>
A:Experimental source: venom

A:Note: sequence extracted from NCBI backbone (NCBIP:135337)

Query Match 14.5%; Score 38; DB 2; Length 30;
Best Local Similarity 42.1%; Pred. No. 1.9e+02;
Matches 8; Conservative 3; Mismatches 6; Indels 2; Gaps 1;

OY 18 PSHMRPLEQ--IRLLTKPE 34
DB 3 PSDMSPEGHGCVFTEPEQ 21

RESULT 6
A05291
C-reactive protein - plaice (fragment)
C:Species: Pleuronectes platessa (plaice)
C:Date: 05-Jun-1987 #sequence_revision 05-Jun-1987 #text_change 22-Apr-1995
C:Accession: A05291
R:Pepys, M.B.; de Beer, F.C.; Milstein, C.P.; March, J.F.; Feinstein, A.; Butress, N.
Biochim. Biophys. Acta 704, 123-133, 1982
A:Reference number: A90642; MUID:8232106
A:Contents: tentative sequence
A:Accession: A05291
A:Molecule type: protein
A:Residues: 1-92 <PEP>
C:Superfamily: C-reactive protein
C:Keywords: acute phase; pentraxin; plasma

Query Match 13.7%; Score 36; DB 2; Length 32;
Best Local Similarity 42.9%; Pred. No. 3.7e+02;
Matches 6; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY 1 MNLVPMNPLVMNGF 14
DB 17 VELIPMKPLNLRAF 30

RESULT 7
E64740
hypothetical protein b0165 - Escherichia coli
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 08-Oct-1999
C:Accession: E64740
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617
A:Accession: E64740
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-43 <BLAT>
A:Cross-references: GB:AE000126; GB:U00096; NID:g1786358; PIDN:AACT3276.1; PID:g17863
A:Experimental source: strain K-12, substrain MG1655

Query Match 13.7%; Score 36; DB 2; Length 43;
Best Local Similarity 46.2%; Pred. No. 5.1e+02;
Matches 6; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 8 PLVNGFCRPSH 20
DB 21 PPIFAGTCRDSH 33

RESULT 8
B96843
protein TELIS.1 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: B96843
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 6, 2002, 09:25:56 ; Search time 12.1 Seconds
(without alignments)
13.018 Million cell updates/sec

Title: US-09-897-042-14

Perfect score: 37

Sequence: 1 MNLVPMN 7

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 35098

Minimum DB seq length: 0
Maximum DB seq length: 7

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : Issued_Patents_AA:*

- 1: /cgn2-6/ptodata/2/1aa/5A.COMB.pep:*
- 2: /cgn2-6/ptodata/2/1aa/5B.COMB.pep:*
- 3: /cgn2-6/ptodata/2/1aa/6A.COMB.pep:*
- 4: /cgn2-6/ptodata/2/1aa/6B.COMB.pep:*
- 5: /cgn2-6/ptodata/2/1aa/PCTUS.COMB.pep:*
- 6: /cgn2-6/ptodata/2/1aa/Backfilest.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	21	56.8	6	2	US-08-910-484-7
2	20	54.1	6	3	US-08-335-844A-67
3	19	51.4	5	4	US-09-236-160-33
4	18	48.6	5	6	5210075-68
5	18	48.6	6	2	US-08-458-109-40
6	18	48.6	6	4	US-08-557-006C-13
7	17	45.9	5	4	US-09-236-160-48
8	17	45.9	6	1	US-08-325-509-46
9	17	45.9	6	1	US-08-595-718A-1
10	17	45.9	6	1	US-08-595-718A-2
11	17	45.9	6	1	US-08-595-718A-3
12	17	45.9	6	1	US-08-595-718A-6
13	17	45.9	6	1	US-08-595-718A-7
14	17	45.9	6	1	US-08-595-718A-8
15	17	45.9	6	1	US-08-447-010-23
16	17	45.9	7	4	US-09-371-814-65
17	17	45.9	7	4	US-09-371-710-22
18	17	45.9	7	4	US-09-230-603-8
19	17	45.9	7	4	US-09-648-386-22
20	16	43.2	4	2	US-08-429-964-41
21	16	43.2	4	4	US-09-053-003-55
22	16	43.2	4	5	PCT-US93-08062-41
23	16	43.2	5	1	US-08-033-873-19
24	16	43.2	5	1	US-08-331-383-32
25	16	43.2	5	1	US-08-549-008-41
26	16	43.2	5	1	US-08-447-010-16
27	16	43.2	5	2	US-08-336-832-19

28	16	43.2	5	3	US-08-802-981-4	Sequence 4, Appli
29	16	43.2	5	4	US-08-988-705-19	Sequence 19, Appli
30	16	43.2	5	4	US-09-236-160-27	Sequence 27, Appli
31	16	43.2	6	1	US-08-414-520A-1	Sequence 1, Appli
32	16	43.2	6	1	US-08-260-199A-19	Sequence 19, Appli
33	16	43.2	6	1	US-08-260-199A-27	Sequence 27, Appli
34	16	43.2	6	1	US-08-260-199A-29	Sequence 29, Appli
35	16	43.2	6	1	US-08-408-604A-45	Sequence 45, Appli
36	16	43.2	6	3	US-08-802-981-187	Sequence 187, App
37	16	43.2	6	5	PCT-US93-09626-45	Sequence 45, Appli
38	16	43.2	6	5	PCT-US93-09626-46	Sequence 46, Appli
39	16	43.2	7	2	US-08-770-301A-7	Sequence 7, Appli
40	16	43.2	7	2	US-08-630-645-21	Sequence 21, Appli
41	16	43.2	7	3	US-09-175-581-7	Sequence 7, Appli
42	16	43.2	7	5	PCT-US96-10220-21	Sequence 21, Appli
43	16	43.2	4	1	US-07-906-349A-16	Sequence 16, Appli
44	15	40.5	4	1	US-08-384-618-6	Sequence 6, Appli
45	15	40.5	4	1	US-08-102-757-23	Sequence 23, Appli
46	15	40.5	4	1	US-08-002-202-7	Sequence 7, Appli
47	15	40.5	4	1	US-08-167-035-16	Sequence 16, Appli
48	15	40.5	4	1	US-08-255-272-21	Sequence 21, Appli
49	15	40.5	4	1	US-08-208-887A-16	Sequence 16, Appli
50	15	40.5	4	1		

ALIGNMENTS

RESULT 1
US-08-910-484-7
Sequence 7, Application US/08910484
Patent No. 5914244
GENERAL INFORMATION:
APPLICANT: Coen, Donald M.
TITLE OF INVENTION: US97 FUSION PROTEINS AND METHODS OF USE
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: FASTSEQ for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/910,484
FILING DATE: 25-JUL-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/022,888
FILING DATE: 25-JUL-1996
ATTORNEY/AGENT INFORMATION:
NAME: Freeman, John W.
REGISTRATION NUMBER: 29,066
REFERENCE/DOCKET NUMBER: 00246/202001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 6 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-910-484-7

Query Match 56.8%; Score 21; DB 2; Length 6;

Best Local Similarity 50.0%; Pred. No. 1.6e+05;
Matches 3; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 2 NLVPMN 7
: : 111
DB 1 DITPMN 6

RESULT 2

US-08-335-844A-67
Sequence 67, Application US/08335844A
Patent No. 6065503

GENERAL INFORMATION:

APPLICANT: GRAHAM, MARGARET
APPLICANT: SMITH, TREVOR STANLEY
APPLICANT: MUNN, EDWARD ALBERT
APPLICANT: KNOX, DAVID PATRICK
APPLICANT: OLIVER, JOANNA JANE
APPLICANT: NEWTON, SUSAN ELIZABETH
TITLE OF INVENTION: RECOMBINANT DNA MOLECULES ENCODING
TITLE OF INVENTION: AMINOPEPTIDASE ENZYMES AND THEIR USE IN THE PREPARATION OF
TITLE OF INVENTION: VACCINES AGAINST HELMINTH INFECTIONS
NUMBER OF SEQUENCES: 73
CORRESPONDENCE ADDRESS:
ADDRESSEE: Rothwell, Flagg, Ernst & Kurtz
STREET: Suite 701-E, 555 Thirteenth St., N.W.
CITY: Washington
STATE: D. C.
COUNTRY: U.S.A.
ZIP: 20004

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/335,844A
FILING DATE: 09-JAN-1995
CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: GB PCT/GB93/00943
FILING DATE: 06-MAY-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9209936
FILING DATE: 08-MAY-1992
ATTORNEY/AGENT INFORMATION:
NAME: WALKER, Barbara W.
REGISTRATION NUMBER: 35,400
REFERENCE/DOCKET NUMBER: 1181-223A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)783-6040
TELEFAX: (202)783-6031

INFORMATION FOR SEQ ID NO: 67:

SEQUENCE CHARACTERISTICS:
LENGTH: 6 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-335-844A-67

Query Match 54.1%; Score 20; DB 3; Length 6;
Best Local Similarity 50.0%; Pred. No. 1.6e+05;
Matches 3; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 NLVPM 6
: : 111
DB 1 NLTP1 6

RESULT 3

US-09-236-160-33
Sequence 33, Application US/09236160A

Patent No. 6307090
GENERAL INFORMATION:
APPLICANT: BURKE JR., Terrence R.
TITLE OF INVENTION: ACYLATED OLIGOPEPTIDE DERIVATIVES HAVING CELL SIGNAL
TITLE OF INVENTION: INHIBITING ACTIVITY
FILE REFERENCE: 175895
CURRENT APPLICATION NUMBER: US/09/236,160A
CURRENT FILING DATE: 1999-01-22
NUMBER OF SEQ ID NOS: 55
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO: 33
LENGTH: 5
TYPE: PPT
ORGANISM: mammalian
US-09-236-160-33

Query Match 51.4%; Score 19; DB 4; Length 5;
Best Local Similarity 75.0%; Pred. No. 1.6e+05;
Matches 3; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 4 VPNN 7
: : 111
DB 1 MPNN 4

RESULT 4

5210075-68
Patent No. 5210075
APPLICANT: SCHOLTZ, WOLFGANG;CHIANG, SHIU-LANG;NACARAJAN,
;GOBILOBI, THOMAS J.
TITLE OF INVENTION: INTERLEUKIN 6 ANTAGONIST PEPTIDES
NUMBER OF SEQUENCES: 69
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/480,868
FILING DATE: 16-FEB-1990
SEQ ID NO: 68
LENGTH: 5
5210075-68

Query Match 48.6%; Score 18; DB 6; Length 5;
Best Local Similarity 75.0%; Pred. No. 1.6e+05;
Matches 3; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 NLVP 5
: : 111
DB 1 NLTP 4

RESULT 5

US-08-458-109-40
Sequence 40, Application US/08458109
Patent No. 5968791
GENERAL INFORMATION:
APPLICANT: Calgene, Inc.
TITLE OF INVENTION: PLANT LYSOPHOSPHATIDIC
TITLE OF INVENTION: ACID ACYLTRANSFERRASES
NUMBER OF SEQUENCES: 44
CORRESPONDENCE ADDRESS:
ADDRESSEE: Calgene, Inc.
STREET: 1920 Fifth Street
CITY: Davis
STATE: CA
COUNTRY: USA
ZIP: 95616

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 2.0 MB
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Macintosh 7.1
SOFTWARE: Microsoft Word 5.1(a)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/458,109

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 6, 2002, 09:18:51 ; Search time 25.55 Seconds

(without alignments)
20.294 Million cell updates/sec

Title: US-09-897-042-14

Perfect score: 37
Sequence: 1 MNLVPMN 7

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 47201

Minimum DB seq length: 0
Maximum DB seq length: 7

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 50 summaries

Database :

A_Geneseq_1101:*

1: /SID2/gcgdata/geneseq/geneseq/AA1980.DAT:*

2: /SID2/gcgdata/geneseq/geneseq/AA1981.DAT:*

3: /SID2/gcgdata/geneseq/geneseq/AA1982.DAT:*

4: /SID2/gcgdata/geneseq/geneseq/AA1983.DAT:*

5: /SID2/gcgdata/geneseq/geneseq/AA1984.DAT:*

6: /SID2/gcgdata/geneseq/geneseq/AA1985.DAT:*

7: /SID2/gcgdata/geneseq/geneseq/AA1986.DAT:*

8: /SID2/gcgdata/geneseq/geneseq/AA1987.DAT:*

9: /SID2/gcgdata/geneseq/geneseq/AA1988.DAT:*

10: /SID2/gcgdata/geneseq/geneseq/AA1989.DAT:*

11: /SID2/gcgdata/geneseq/geneseq/AA1990.DAT:*

12: /SID2/gcgdata/geneseq/geneseq/AA1991.DAT:*

13: /SID2/gcgdata/geneseq/geneseq/AA1992.DAT:*

14: /SID2/gcgdata/geneseq/geneseq/AA1993.DAT:*

15: /SID2/gcgdata/geneseq/geneseq/AA1994.DAT:*

16: /SID2/gcgdata/geneseq/geneseq/AA1995.DAT:*

17: /SID2/gcgdata/geneseq/geneseq/AA1996.DAT:*

18: /SID2/gcgdata/geneseq/geneseq/AA1997.DAT:*

19: /SID2/gcgdata/geneseq/geneseq/AA1998.DAT:*

20: /SID2/gcgdata/geneseq/geneseq/AA1999.DAT:*

21: /SID2/gcgdata/geneseq/geneseq/AA2000.DAT:*

22: /SID2/gcgdata/geneseq/geneseq/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	56.8	5	19	AAW35997	Human myocardium t
2	54.1	6	19	AAW75338	Hexapeptide #3 bin
3	54.1	6	19	AAW75270	Hexapeptide #3 bin
4	54.1	7	22	AAW47356	Cell lysis fractio
5	48.6	6	16	AAW87710	Peptide #9 in nove
6	48.6	6	16	AAW64314	Internal sequence
7	48.6	6	21	AAW69214	Peptide fragment (r
8	48.6	7	12	AAW10518	Fusinus inhibitory
9	48.6	7	19	AAW81797	Seq ID 18 from Del
10	48.6	7	20	AAW01780	Active fragment of
11	45.9	5	18	AAW22532	Peptide #6 having

12	17	45.9	6	14	AAW30840	Sequence of peptid
13	17	45.9	6	18	AAW22527	Peptide #1 having
14	17	45.9	6	18	AAW22528	Peptide #2 having
15	17	45.9	6	18	AAW22529	Peptide #3 having
16	17	45.9	6	18	AAW22535	Peptide #9 tested
17	17	45.9	6	18	AAW22536	Peptide #10 tested
18	17	45.9	6	18	AAW22534	Peptide #8 tested
19	17	45.9	6	18	AAW72228	Human MHC Class II
20	17	45.9	6	21	AAW3606	Aspergillus niger
21	17	45.9	6	21	AAW3608	Aspergillus niger
22	17	45.9	6	21	AAW3610	Aspergillus niger
23	17	45.9	7	18	AAW15795	Adherent factor p
24	17	45.9	7	19	AAW50220	Recombinant beta-g
25	17	45.9	7	22	AAW49427	Peptide #6 used to
26	16	43.2	4	12	AAW15772	Farnesyl-protein t
27	16	43.2	4	15	AAW49769	Farnesyl transfera
28	16	43.2	4	16	AAW77833	Farnesyl transfera
29	16	43.2	4	17	AAW04459	Sequence of peptid
30	16	43.2	5	14	AAW30834	Fluorogenic protea
31	16	43.2	5	19	AAW82095	Peptide containing
32	16	43.2	5	19	AAW46560	Alpha-1 antitrypsi
33	16	43.2	5	22	AAW73337	Streptokinase-bind
34	16	43.2	6	16	AAW75677	Peptide against PI
35	16	43.2	6	17	AAW07284	Mab PK99H/PAK pept
36	16	43.2	6	17	AAW97583	Mab PK99H/PAK pept
37	16	43.2	6	17	AAW97591	Mab PK99H/PAK pept
38	16	43.2	6	17	AAW97593	Fluorogenic protea
39	16	43.2	6	19	AAW82039	SH2 domain binding
40	16	43.2	6	19	AAW78565	SH2 domain binding
41	16	43.2	6	19	AAW78566	SH2 domain binding
42	16	43.2	6	20	AAW03909	Hepatitis C inhibi
43	16	43.2	6	20	AAW90096	Rat CARP (62-102)
44	16	43.2	6	21	AAW24302	Prostate tumour as
45	16	43.2	6	21	AAW26204	Protein-tyrosine k
46	16	43.2	6	22	AAW82152	Signal transductio
47	16	43.2	6	22	AAW82153	Hypotensive peptid
48	16	43.2	7	11	AAW03864	Streptokinase-bind
49	16	43.2	7	16	AAW75676	Hepatitis C virus
50	16	43.2	7	16	AAW84394	Hepatitis C virus

ALIGNMENTS

RESULT 1

AAW35997

ID AAW35997 standard; peptide: 5 AA.

AC AAW35997:

DT 05-MAR-1998 (first entry)

XX Human myocardium troponin T epitope.

DE Human; myocardium; troponin T; monoclonal antibody; epitope;

KW skeletal muscle; immunoassay.

KM Homo sapiens.

OS

XX JP09271392-A.

FN

XX 21-OCT-1997.

PD

XX 09-APR-1996; 96JP-0086932.

PF

XX 09-APR-1996; 96JP-0086932.

PR

XX (TERU) TERUMO CORP.

PA WPI: 1998-003027/01.

XX Anti-human myocardium troponin T monoclonal antibody - useful in

PT immunoassay for distinguishing myocardial from skeletal muscle

PT tropoinin T
XX
PS Claim 1; Page 1; 6pp; Japanese.
XX
CC The present sequence represents a peptide epitope of human myocardium
CC tropoinin T, which is recognised by a new anti-human myocardium tropoinin
CC T monoclonal antibody. In an example from the specification, a hybridoma
CC producing anti-human myocardium tropoinin T IgG monoclonal antibody was
CC screened. An antibody used in human myocardium tropoinin T measuring
CC system was selected. The epitope of human myocardium tropoinin T-specific
CC antibody was analysed. An enzymatic immunoassay for tropoinin T
CC measurement was carried out. The measurements in a chronic renal disease
CC patient and a diabetes patient of normal heart function were examined.
CC The effect of co-presence of skeletal muscle tropoinin T was evaluated.
CC The anti-human myocardium tropoinin T monoclonal antibody is used in an
CC immunoassay for the determination of tropoinin T in a solution. The
CC method can determine tropoinin T exactly with no effect of skeletal
CC muscle tropoinin T.
Sequence 5 AA:

Query Match 56.8%; Score 21; DB 19; Length 5;
Best Local Similarity 100.0%; Pred. No. 4.3e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 NLVP 5
| | | |
Db 2 nlvp 5

RESULT 2
AAW75338
ID AAW75338 standard; peptide: 6 AA.

AC AAW75338;

DT 02-FEB-1999 (first entry)

DE Hexapeptide #3 binds anti-Ad5 fibre head MAb 1D6.3.

XX Cellular receptor; virus: immobilisation; monoclonal antibody; fibre;
KW viral surface protein; hexapeptide expression library; adenovirus;
KW major histocompatibility complex; MHC; fibronectin; gene therapy;
XX genetic disease; acquired immune deficiency syndrome; AIDS; cancer.

OS Synthetic.

PN FR2758821-A1.

PD 31-JUL-1998.

PE 30-JAN-1997; 97FR-0001005.

PR 30-JAN-1997; 97FR-0001005.

PR 09-SEP-1997; 97FR-0011166.

PA (CNRS) CNRS CENT NAT RECH SCI.

PI Boulanger P, Hong SS, Karayan L;

PT WPI; 1998-416458/36.

XX Use of peptide(s) for facilitating or modulating attachment of
PT adenovirus to cells - useful for, e.g. treating or preventing
PT infection and improving uptake of gene therapy vectors
XX

PS Example 2; Fig 1; 48pp; French.

XX The invention relates to methods for selecting and identifying a
CC cellular receptor for a virus, by immobilising, on a support, a
CC monoclonal antibody targeted to a viral surface protein that determines
CC attachment of the virus to the receptor. The immobilised antibody is

CC incubated with a hexapeptide expression library and peptides bound to the
CC immobilised antibody are eluted by competitive binding with recombinant
CC fragments of the viral surface protein. In a reverse method, the viral
CC surface protein is immobilised and incubated with the peptide library. In
CC this case, the bound peptides are eluted by competitive binding with the
CC monoclonal antibody. The hexapeptides AAW75336-W75345 represent
CC peptides isolated by biopanning the library with the immobilised
CC monoclonal antibody (MAb) 1D6.3 and eluted using an adenovirus type 5
CC fibre head protein (AAW75334).
CC The methods are used to identify peptides from MHC Class I and/or
CC type III fibronectin proteins that allow or facilitate attachment by
CC adenovirus (Ad) to host cells and/or entry into the cells, and to
CC identify ligands that modulate Ad infection mediated by these peptides,
CC e.g. to treat or prevent Ad infections or to facilitate infection by Ad
CC gene therapy vectors used to treat genetic diseases, acquired immune
CC deficiency syndrome or cancer.
Sequence 6 AA:

Query Match 54.1%; Score 20; DB 19; Length 6;
Best Local Similarity 60.0%; Pred. No. 4.3e+05;
Matches 3; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 3 LVPWN 7
| | | |
Db 1 lvpfn 5

RESULT 3
AAW75270
ID AAW75270 standard; peptide: 6 AA.

AC AAW75270;

DT 02-FEB-1999 (first entry)

DE Hexapeptide #3 binds anti-Ad5 fibre head MAb 1D6.3.

XX Cellular receptor; virus: immobilisation; monoclonal antibody; fibre;
KW viral surface protein; hexapeptide expression library; adenovirus;
KW major histocompatibility complex; MHC; fibronectin; gene therapy;
XX genetic disease; acquired immune deficiency syndrome; AIDS; cancer.

OS Synthetic.

PN FR2758885-A1.

PD 31-JUL-1998.

PE 30-JAN-1997; 97FR-0001005.

PR 30-JAN-1997; 97FR-0001005.

PR 10-JUL-1997; 97FR-0008796.

PA (CNRS) CNRS CENT NAT RECH SCI.

PI Boulanger P, Hong SS, Karayan L;

PT WPI; 1998-416493/36.

XX Selection and identification of cellular receptors for viruses -
PT used to control attachment and entry of adenovirus into cells, e.g.
PT for treating infection or in gene therapy
XX

PS Example 2; Fig 1; 43pp; French.

XX The invention relates to methods for selecting and identifying a
CC cellular receptor for a virus, by immobilising, on a support, a
CC monoclonal antibody targeted to a viral surface protein that determines
CC attachment of the virus to the receptor. The immobilised antibody is
CC incubated with a hexapeptide expression library and peptides bound to the
CC immobilised antibody are eluted by competitive binding with recombinant

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 6, 2002, 09:35:21 ; Search time 24.53 Seconds
(without alignments)
41.741 Million cell updates/sec

Title: US-09-897-042-14

Perfect score: 37

Sequence: 1 MNLVPMN 7

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 segs, 146272329 residues

Total number of hits satisfying chosen parameters: 65

Minimum DB seq length: 0
Maximum DB seq length: 7

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database :

1: SPTRMBL_17:*
2: sp_archaea:*
3: sp_bacteria:*
4: sp_fungi:*
5: sp_human:*
6: sp_invertebrate:*
7: sp_mammal:*
8: sp_mmc:*
9: sp_organelle:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

No.	Score	Query Match	Length	DB ID	Description
1	13	35.1	7	2	050556 actinobacil
2	13	35.1	7	13	042564 fugu rubrip
3	12	32.4	5	13	P82099 litorea rub
4	12	32.4	6	13	P82096 litorea rub
5	10	27.0	7	8	098866 spincia ol
6	10	27.0	7	12	007624 rous sarcom
7	10	27.0	7	13	P82065 litorea rub
8	9	24.3	7	2	007354 synechococ
9	9	24.3	7	2	P72081 nocardia la
10	9	24.3	7	10	P93233 lycopersico
11	9	24.3	7	12	066205 porcine tra
12	8	21.6	4	11	008433 rattus norv
13	8	21.6	7	2	054248 streptomyc
14	8	21.6	7	10	09C5B3 arabidopsis
15	8	21.6	7	11	063668 rattus norv
16	8	21.6	7	12	066113 cherry leaf
17	8	21.6	7	12	067113 influenza a
18	7	18.9	7	2	034028 pseudomonas
19	7	18.9	7	8	P92421 psathyrosta

20	7	18.9	7	8	P92385 hordeum mat
21	7	18.9	7	8	P92210 agropyron c
22	7	18.9	7	8	P92214 amblyopyrum
23	7	18.9	7	8	P92218 australopyr
24	7	18.9	7	8	P92221 bromus iner
25	7	18.9	7	8	P92326 crithopsis
26	7	18.9	7	8	P92372 haynaldia v
27	7	18.9	7	8	P92381 hordeum bra
28	7	18.9	7	8	P92387 heteranthel
29	7	18.9	7	8	P92390 hordeum vul
30	7	18.9	7	8	P92393 hordeum vul
31	7	18.9	7	8	P92425 pseudoregn
32	7	18.9	7	8	P92427 peridiclyon
33	7	18.9	7	8	P92430 aegilops ta
34	7	18.9	7	8	P92442 taenialtheru
35	7	18.9	7	8	P92440 thnopyrum
36	7	18.9	7	8	P92403 lophopyrum
37	7	18.9	7	13	P82101 litorea rub
38	6	16.2	5	10	099007 hordeum vul
39	6	16.2	7	2	047505 escherichia
40	6	16.2	7	6	028742 oryctolagus
41	6	16.2	7	8	099182 gnatholebia
42	6	16.2	7	12	065578 bovine herp
43	5	13.5	6	4	008720 homo sapien
44	5	13.5	6	10	P82541 spinacia ol
45	5	13.5	7	2	047029 enterobacte
46	5	13.5	7	11	063480 rattus norv
47	5	13.5	7	12	09Y010 porcine tra
48	4	10.8	5	13	P82070 litorea rub
49	4	10.8	5	13	P82073 litorea rub
50	4	10.8	5	13	P82100 litorea rub

ALIGNMENTS

RESULT	1	PRELIMINARY:	PRT:	7 AA.
ID	050556			
AC	050556			
DT	01-JUN-1998 (TREMUREL.06, Created)			
DT	01-JUN-1998 (TREMUREL.06, Last sequence update)			
DT	01-JUN-2001 (TREMUREL.17, Last annotation update)			
DE	GLYA (FRAGMENT).			
GN	GLYA.			
OS	Actinobacillus actinomycetemcomitans (Haemophilus			
OC	actinomycetemcomitans)			
OC	Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;			
OC	Actinobacillus.			
ON	NCBI_TaxID=714;			
OX	11			
RP	SEQUENCE FROM N.A.			
RC	SPRAIN=ATCC 33384;			
RX	MEDLINE=96355846; PubMed=8751884;			
RA	Kolodrubetz D., Spltznagel J. Jr., Wang B., Phillips L.H., Jacobs C.,			
RT	Kraig E.;			
RT	"cis Elements and trans factors are both important in strain-specific			
RT	regulation of the leukotoxin gene in Actinobacillus			
RT	actinomycetemcomitans.";			
RL	Infect. Immun. 64:3451-3460(1996).			
DR	EMBL; U51862; AAB88721.1; -.			
FT	NON_TER			
SO	SEQUENCE	7 AA; 832 MW; 6DCBA2D767340420 CRC64;		

Query Match 35.1%; Score 13; DB 2; Length 7;
Best Local Similarity 40.0%; Pred. NO. 4.7e+05;
Matches 2; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 MNLVPM 6
| : : : :
Db 1 NRLPV 5

RESULT 2
ID 042564 PRELIMINARY: PRT: 7 AA.
AC 042564.
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
DE TRUNCATED VOLTAGE-GATED SODIUM CHANNEL ALPHA SUBUNIT (FRAGMENT).
GN SCNA.
OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Acanthomorphi; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontidae; Takifugu.
OX NCBI_TaxID=31033;
RN [1]
RP SEQUENCE FROM N.A.
MEDLINE=97442476; PubMed=9295353;
Plummer N.W., McBurney M.W., Meisler M.H.;
"Alternative splicing of the sodium channel SCNA predicts a truncated
two-domain protein in fetal brain and non-neuronal cells."
RL J. Biol. Chem. 272:24008-24015(1997).
DR EMBL: U97673; AAB80916.1; -.
KW Ionic channel.
FT NON_TER
SQ SEQUENCE 7 AA; 730 MW; 75B72EA2C73772A0 CRC64;

Query Match 35.1%; Score 13; DB 13; Length 7;
Best Local Similarity 66.7%; Pred. No. 4.7e+05;
Matches 2; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 4 VPM 6
DB 1 VPL 3

RESULT 3
ID P82099 PRELIMINARY: PRT: 5 AA.
AC P82099.
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
DE ELECTRIN 3.
OS Litoria rubella (Desert tree frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Hylidae;
OC Litoria.
OX NCBI_TaxID=104895;
RN [1]
RP SEQUENCE.
RC TISSUE=SKIN SECRETION;
RA Wadnitz P.A., Bowle J.H., Tyler M.J., Wallace J.C.;
"Peptides from the skin glands of the Australian buzzing tree frog
RT Litoria electrica. Comparison with the skin peptides from Litoria
RT rubella."
RL Aust. J. Chem. 52:0-0(1999).
KW Amphibian skin; Amidation.
FT MOD_RES
SQ SEQUENCE 5 AA; 630 MW; 66B761FC9A0000 CRC64;

Query Match 33.4%; Score 12; DB 13; Length 5;
Best Local Similarity 100.0%; Pred. No. 4.7e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 PM 6
DB 4 PM 5

RESULT 4
ID P82096 PRELIMINARY: PRT: 6 AA.
AC P82096.
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
DE ELECTRIN 1.
OS Litoria rubella (Desert tree frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Hylidae;
OC Litoria.
OX NCBI_TaxID=104895;
RN [1]
RP SEQUENCE.
RC TISSUE=SKIN SECRETION;
RA Wadnitz P.A., Bowle J.H., Tyler M.J., Wallace J.C.;
"Peptides from the skin glands of the Australian buzzing tree frog
RT Litoria electrica. Comparison with the skin peptides from Litoria
RT rubella."
RL Aust. J. Chem. 52:0-0(1999).
KW Amphibian skin; Amidation.
FT MOD_RES
SQ SEQUENCE 6 AA; 792 MW; 66B3704772C9A000 CRC64;

Query Match 32.4%; Score 12; DB 13; Length 6;
Best Local Similarity 66.7%; Pred. No. 4.7e+05;
Matches 2; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 4 VPM 6
DB 2 VPI 4

RESULT 5
ID 098866 PRELIMINARY: PRT: 7 AA.
AC 098866.
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-MAY-1999 (TREMBlrel. 10, Last annotation update)
DE CYTOCHROME B/F SUBUNIT IV (FRAGMENT).
OS Spinacia oleracea (Spinach).
OC Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Caryophyllidae; Caryophyllales; Chenopodiaceae; Spinacia.
OX NCBI_TaxID=3562;
RN [1]
RP SEQUENCE FROM N.A.
MEDLINE=66120353; PubMed=3003688;
RX Sjibben-Mueller G., Hallick R.B., Alt J., Westhoff P., Herrmann R.G.;
"Spinach plastid genes coding for initiation factor IF-1, ribosomal
RT protein S11 and RNA polymerase alpha-subunit."
RL Nucleic Acids Res. 14:1029-1044(1986).
DR EMBL: X03496; CAA27215.1; -.
KW Chloroplast.
FT NON_TER
SQ SEQUENCE 7 AA; 907 MW; 644729D77409C420 CRC64;

Query Match 27.0%; Score 10; DB 8; Length 7;
Best Local Similarity 50.0%; Pred. No. 4.7e+05;
Matches 2; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 NLVP 5
DB 1 NFRP 4

RESULT 6
007624

OM protein - protein search, using sw model

(without alignments)
26.136 Million cell updates/sec

Sequence: 1 MNLVPMN 7

[illegible]

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 74

Maximum	DB	seq	length:	0
Maximum	DB	seq	length:	7

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 50 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query	Score	Match	Length	DB	ID	Description
No.	No.						
1	35.1	7	1	CARP_MYTD	P10420	mytilus edulis	
2	35.1	7	1	MYO_AMLCA	P15513	aplysia catenella	
3	32.4	6	1	C1P1_MYTD	P13736	mytilus edulis	
4	32.4	6	1	C1P2_MYTD	P13737	mytilus edulis	
5	32.4	6	1	TM0E_SARBU	P14395	sarcophaga	
6	27.0	6	1	VP19_HSV1K	P23210	herpes simplex	
7	27.0	7	1	FARB_CALVO	P41866	calliphora	
8	27.0	4	1	UN06_PINPS	P81575	pinus pinaster	
9	24.3	4	1	RM01_YEAST	P36515	saccharomyces	
10	24.3	7	1	CCP1_ENTFA	P20104	enterococcus	
11	24.3	7	1	FAR4_PANRE	P41875	panagrellus	
12	24.3	7	1	MRP1_LEPDE	P42984	leptinotarsus	
13	21.6	5	1	MSCL_SALTY	P39446	salmonella	
14	21.6	5	1	PRCT_PERAM	P01373	periplaneta	
15	21.6	3	1	TRP1_PSEPU	P36414	pseudomonas	
16	18.9	6	1	THVL_PIG	P01151	sus scrofa	
17	18.9	4	1	DCML_PSECH	P19916	pseudomonas	
18	18.9	4	1	TUFT_HUMAN	P19516	pseudomonas	
19	18.9	5	1	B10B_SALTY	P01858	homo sapiens	
20	18.9	5	1	BPP7_BOTIN	P12678	salmonella	
21	18.9	5	1	PAP2_PARMA	P30425	bothriopsis	
22	18.9	5	1	SUGA_ACHDO	P81864	parachirius	
23	18.9	6	1	OVW_LEPDE	P19991	acheta domestica	
24	18.9	7	1	ALL3_CARMA	P42985	leptinotarsus	
25	18.9	7	1	ALL4_CARMA	P81806	carcinus maenas	
26	18.9	7	1	ALL5_CARMA	P81807	carcinus maenas	
27	18.9	7	1	CHOX_ALCSP	P81808	carcinus maenas	
28	18.9	7	1	FAR1_HELTI	P16101	alcaligenes	
29	18.9	7	1	FAR2_PROCL	P41871	heliopsis	
30	18.9	7	1	FAR3_PROCL	P38495	procambarus	
31	18.9	7	1	GFRP_MOUSE	P38498	procambarus	
32	18.9	7	1	LANC_CARUI	P90925	mus musculus	
33	18.9	7	1	UF04_MOUSE	P36360	carobacter	
						mus musculus	

ALIGNMENTS

34	7	1	MMAL_ACHVU	P35919	achatina fur
35	7	1	MMAL2_ACHVU	P35920	achatina fur
36	7	1	MMAL3_ACHVU	P35921	achatina fur
37	6	5	TPIS_CANPA	P34714	canis famili
38	6	5	TRM3_ECOLI	P31973	escherichia
39	6	1	ASP2_LACNS	P34965	lactobacilli
40	6	6	FARP_MONEX	P19566	montezia ex
41	6	6	UN06_CLOPA	P13151	clostridium
42	6	7	FAR1_ASCSU	P11889	ascaris suu
43	6	7	ICAO_DACDE	P06294	dactylium d
44	5	4	DCWS_PSECH	P191918	pseudomonas
45	5	4	FAR4_HIRME	P42563	hirudo medi
46	5	4	FMRF_MACNI	P11012	macrobacter
47	5	5	BIOA_CITFR	P13071	citrobacter
48	5	5	BIOA_SALRY	P12657	salmonella
49	5	5	BIOB_CITFR	P12997	cytochrome
50	5	6	ACPH_NABIT	P25154	oryctolagus

RESULT 1

ID	CARP_MYTEID	STANDARD:	PRT:	7 AA.
AC	P10420:			
DT	01-MAR-1989 (Rel. 10, Created)			
DT	01-MAR-1989 (Rel. 10, Last sequence update)			
DT	01-MAR-1989 (Rel. 10, Last annotation update)			
DE	CATCH-RELAXING PEPTIDE (CARP).			
OS	Mytilus edulis (Blue mussel).			
OC	Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloida;			
OC	Mytiloidea; Mytilidae; Mytilus.			
OX	NCBI_TaxID=6550;			
RN	[1]			
RP	SEQUENCE.			
RA	MEDLINE=88052022; PubMed=3676797;			
RA	Hirata T., Kubota I., Takabatake I., Kawahara A., Shimamoto N.,			
RA	Muneoka Y.;			
RT	"Catch-relaxing peptide isolated from Mytilus pedal ganglia.";			
RL	Brain Res. 422:374-376(1987).			
CC	-!- FUNCTION: THIS PEPTIDE EXHIBITS BOTH POTENTIATING (CONTRACTORY			
CC	AND INHIBITORY (RELAXATION) EFFECTS ON THE ANTERIOR BYSSUS			
CC	RETRACTOR MUSCLE.			
CC	PIR: A29342; ECDMCR.			
DR	PIR: A29342; ECDMCR.			
KM	Hormone; Amidation.			
FT	MOD_RES			
SO	SEQUENCE	7 AA;	831 MW;	67340726876699B0 CRC64;

Query Match	35.18;	Score 13;	DB 1;	Length 7;
Best Local Similarity	66.78;	Prod NO	10+05;	

Matches 2; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 4 VPM 6

Db 2 MPM 4

RESULT 2

ID	MYOM	APCA	STANDARD;	PRT;	7 AA.
AC	P15513				
DT	01-APR-1990	(Rel. 14,	Created)		
DT	01-APR-1990	(Rel. 14,	Last sequence update)		
DT	15-JUL-1998	(Rel. 36,	Last annotation update)		
DE	MYOMODULIN (NEURON B16 PEPTIDE) .				
OS	Aplysia californica (California sea hare).				
OC	Eukaryota; Metazoa; Mollusca; Gastropoda; Opisthobranchia; Anaspidaea				
AP	Aplysiidae; Aplysia.				
OX	NCBI_TaxId=6500;				
RN	[1]				
RP	SEQUENCE.				

RC TISSUE-Buccal muscle;
RX MEDLINE-87261010: PubMed-3474664;
RA "Cropper E.C., Tenenbaum R., Kolks M.A.G., Kupfermann I., Weiss K.R.;
RT "Myomodulin: a bioactive neuropeptide present in an identified
RT cholinergic buccal motor neuron of Aplysia."
RL Proc. Natl. Acad. Sci. U.S.A. 84:5483-5486(1987).
CC -1- FUNCTION: EXOGENOUS APPLICATION OF THE PEPTIDE POTENTIATES ARC
CC MUSCLE CONTRACTIONS.
DR PIR: A28340; A28340.
KW Neuropeptide; Amidation.
FT MOD_RES 7
SQ SEQUENCE 7 AA: 847 MW: 6734072685B8700 CRC64;

Query Match 35.1%; Score 13; DB 1; Length 7;
Best Local Similarity 66.7%; Pred. No. 1e+05;
Matches 2; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 5 PM 7
DB 1 PMS 3

RESULT 3
CIP1_MYTED
ID CIP1_MYTED STANDARD: PRT: 6 AA.
AC P13736;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE CONTRACTION-INHIBITING PEPTIDE I (MIP I).
OS Mytilus edulis (Blue mussel).
CC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloidea;
OC Mytiloidea; Mytilidae; Mytilus.
OX NCBI_Taxid=6550;
RN [1]
RP SEQUENCE.
RC TISSUE-Pedal ganglion;
RX MEDLINE-88240357; PubMed-3377776;
RA Hirata T., Kubota I., Iwasawa N., Takabatake I., Ikeda T., Muneoka Y.;
RT "Structures and actions of Mytilus inhibitory peptides."
RL Blochem. Biophys. Res. Commun. 152:1376-1382(1988).
CC -1- FUNCTION: INHIBITORY ACTION ON CONTRACTIONS IN SEVERAL MOLLUSCAN
CC MUSCLES.
CC -1- SIMILARITY: TO MIP II.
DR PIR: A27696; A27696.
KW Hormone; Amidation.
FT MOD_RES 6
SQ SEQUENCE 6 AA: 637 MW: 72C9C68775B81000 CRC64;

Query Match 32.4%; Score 12; DB 1; Length 6;
Best Local Similarity 100.0%; Pred. No. 1e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 PM 6
DB 3 PM 4

RESULT 4
CIP2_MYTED
ID CIP2_MYTED STANDARD: PRT: 6 AA.
AC P13737;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE CONTRACTION-INHIBITING PEPTIDE II (MIP II).
OS Mytilus edulis (Blue mussel).
CC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloidea;
OC Mytiloidea; Mytilidae; Mytilus.
OX NCBI_Taxid=6550;
RN [1]

RP SEQUENCE.
RC TISSUE-Pedal ganglion;
RX MEDLINE-88240357; PubMed-3377776;
RA Hirata T., Kubota I., Iwasawa N., Takabatake I., Ikeda T., Muneoka Y.;
RT "Structures and actions of Mytilus inhibitory peptides."
RL Blochem. Biophys. Res. Commun. 152:1376-1382(1988).
CC -1- FUNCTION: INHIBITORY ACTION ON CONTRACTIONS IN SEVERAL MOLLUSCAN
CC MUSCLES.
CC -1- SIMILARITY: TO MIP I.
DR PIR: B27696; B27696.
KW Hormone; Amidation.
FT MOD_RES 6
SQ SEQUENCE 6 AA: 621 MW: 72C9C6876DB81000 CRC64;

Query Match 32.4%; Score 12; DB 1; Length 6;
Best Local Similarity 100.0%; Pred. No. 1e+05;
Matches 2; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 PM 6
DB 3 PM 4

RESULT 5
TMOF_SARBU
ID TMOF_SARBU STANDARD: PRT: 6 AA.
AC P41495;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE TRYPSIN-MODULATING OOSTATIC FACTOR (TMOF).
OS Sarcophaga bullata (Grey flesh fly) (Neobellieria bullata).
CC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OX Oestrioidea; Sarcophagidae; Sarcophaga.
RN [1]
RP SEQUENCE, AND SYNTHESIS.
RC TISSUE-Ovary;
RX MEDLINE-94211930; PubMed-8159607;
RA Bylemans D., Borovsky D., Hunt D.F., Shabanowitz J., Grauwels L.,
RA de Loof A.;
RT "Sequencing and characterization of trypsin modulating oostatic
RT factor (TMOF) from the ovaries of the grey fleshfly, Neobellieria
RT (Sarcophaga) bullata."
RL Regul. Pept. 50:61-72(1994).
CC -1- FUNCTION: HAS AN OOSTATIC ACTIVITY. INHIBITS TRYPSIN BIOSYNTHESIS
CC IN THE MIDGUT WHICH INDIRECTLY REDUCES THE VITELLOGENIN
CC CONCENTRATION IN THE HEMOLYMPH RESULTING IN INHIBITION OF OOCYTE
CC DEVELOPMENT.
CC -1- DEVELOPMENTAL STAGE: SYNTHESIZED AND RELEASED FROM FOLLICULAR
CC EPITHELIUM AFTER A BLOOD MEAL.
KW Hormone.
FT MOD_RES 6
SQ SEQUENCE 6 AA: 695 MW: 61E72451B7642000 CRC64;

Query Match 32.4%; Score 12; DB 1; Length 6;
Best Local Similarity 66.7%; Pred. No. 1e+05;
Matches 2; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 5 PM 7
DB 2 PM 4

RESULT 6
VP19_HSVIK
ID VP19_HSVIK STANDARD: PRT: 6 AA.
AC P23210;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 01-DEC-1992 (Rel. 24, Last annotation update)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 6, 2002, 09:30:51 ; Search time 18.86 Seconds
(without alignments)
28.273 Million cell updates/sec

Title: US-09-897-042-14

Perfect score: 37

Sequence: 1 MNLVPMN 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 455

Minimum DB seq length: 0

Maximum DB seq length: 7

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 50 summaries

Database :

1: PIR:68:*
2: PIR1:*
3: PIR2:*
4: PIR3:*
5: PIR4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	16	43.2	7	2	148086
2	13	35.1	4	2	146627
3	13	35.1	6	2	165546
4	13	35.1	7	2	ECMUOK
5	13	35.1	7	2	A28340
6	12	32.4	6	2	A27696
7	12	32.4	6	2	B27696
8	11	29.7	4	2	A61300
9	11	29.7	4	2	S43014
10	11	29.7	4	2	IS4357
11	11	29.7	5	2	A60521
12	11	29.7	5	2	B25655
13	11	29.7	5	2	P00009
14	11	29.7	7	2	P00663
15	11	29.7	7	2	SI7976
16	11	29.7	7	2	E61491
17	11	29.7	7	2	S78024
18	10	27.0	6	2	I49424
19	10	27.0	7	2	S71867
20	10	27.0	7	2	A61081
21	10	27.0	7	2	PC2370
22	10	27.0	7	2	PN0649
23	10	27.0	7	2	PN0150
24	10	27.0	7	2	B44787
25	10	27.0	7	2	C56793
26	10	27.0	7	2	S58797
27	10	27.0	7	2	S45648
28	9	24.3	3	2	T13892
29	9	24.3	4	2	A32039

30	9	24.3	4	2	S17255	ribosomal protein
31	9	24.3	5	2	E42364	flagellar protein
32	9	24.3	5	2	S11127	phosphoprotein, bo
33	9	24.3	6	4	A35039	hypothetical colla
34	9	24.3	6	4	S15596	orf 3 rara 5'-regi
35	9	24.3	7	2	A30812	sex pheromone cCFL
36	9	24.3	7	2	A11483	aspartate transami
37	9	24.3	7	2	I48105	dihydrofolate redu
38	8	21.6	3	2	P00010	angiotensin-conver
39	8	21.6	3	2	I78890	tyrosine protein k
40	8	21.6	4	2	B43848	cell surface adhes
41	8	21.6	4	2	A40135	branched-chain-aml
42	8	21.6	5	1	H0R0HA	proctolin - Americ
43	8	21.6	5	2	JN0862	peptidyl-dipeptida
44	8	21.6	5	2	JN0860	peptidyl-dipeptida
45	8	21.6	5	2	C41225	copper resistance
46	8	21.6	5	2	E60274	major protein anti
47	8	21.6	5	2	B37968	acid proteinase II
48	8	21.6	5	2	A37114	hypoxanthine phosph
49	8	21.6	5	2	A60411	proctolin - Atlant
50	8	21.6	5	2	S53595	hypothetical prote

ALIGNMENTS

RESULT 1

148086 DNA topoisomerase II alpha - Chinese hamster (fragment)

C:Species: Citicellus griseus (Chinese hamster)

C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 05-Nov-1999

C:Accession: 148086

R:Ng, S.W.; Eder, J.P.; Schnipper, L.E.; Chan, V.T.W.

J. Biol. Chem. 270, 25850-25858, 1995

A:Title: Molecular cloning and characterization of the promoter for the Chinese hamst

A:Reference number: 148086, MUID:96029684

A:Accession: 148086

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-7 <RES>

A:Cross-references: EMBL:U34196; NID:91041231; PIDN:MAC52315.1; PID:91041232

Query Match 43.2%, Score 16; DB 2; Length 7;

Best Local Similarity 50.0%; Pred. No. 2.2e+05;

Matches 3; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Oy 1 MNLVPM 6
Db 1 MELSP 6

RESULT 2

146627 hypothetical protein c4 - loblolly pine

C:Species: Pinus taeda (loblolly pine)

C:Date: 18-Feb-2000 #sequence_revision 18-Feb-2000 #text_change 18-Feb-2000

C:Accession: 146627

R:Chang, S.; Puryear, J.; Funkhouser, E.A.; Newton, R.J.; Cairney, J.

submitted to the EMBL Data Library, July 1995

A:Description: Cloning of a chitinase homolog which lacks chitin binding sites and is

A:Reference number: 223105

A:Accession: 146627

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-4 <CH>

A:Cross-references: EMBL:U31309; NID:9974285; PID:9974292

A:Experimental source: strain s6PTx66PT3; 8 month seedlings

Query Match 35.1%, Score 13; DB 2; Length 4;
Best Local Similarity 75.0%; Pred. No. 2.2e+05;
Matches 3; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 MNLV 4
1 1 1
Db 1 MKLV 4

RESULT 3

165546
MHC H2-L antigen - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change 05-Nov-1999
C:Accession: 165546
R:Kimura, A.; Israel, A.; Le Bail, O.; Kourilsky, P.
Cell 44, 261-272, 1986
A:Title: Detailed analysis of the mouse H-2Kb promoter: Enhancer-like sequences and the
A:Reference number: 152778; MUID:86106202
A:Accession: 165546
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-6 <RES>
A:Cross-references: GB:M12483; NID:g199565; PIDN:AAA39663.1; PID:g554234

Query Match 35.1%; Score 13; DB 2; Length 6;
Best Local Similarity 66.7%; Pred. No. 2.2e+05;
Matches 2: Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 3 LVP 5
1 1 1
Db 1 MVP 3

RESULT 4

ECMUCR
catch-relaxing peptide - blue mussel
N:Alternate names: CARP
C:Species: Mytilus edulis (blue mussel)
C:Date: 20-Jun-2000 #sequence_revision 20-Jun-2000 #text_change 20-Jun-2000
C:Accession: A29342
R:Hirata, T.; Kubota, I.; Takabatake, I.; Kawahara, A.; Shimamoto, N.; Muneoka, Y.
Brain Res. 422, 374-376, 1987
A:Title: Catch-relaxing peptide isolated from Mytilus pedal ganglia.
A:Reference number: A29342; MUID:86052022
A:Accession: A29342
A:Molecule type: protein
A:Residues: 1-7 <HIR>
C:Comment: This peptide exhibits both potentiating (contraction) and inhibitory (relaxa
C:Superfamily: Unassigned animal peptides
C:Keywords: amidated carboxyl end; hormone; retractor muscle
F:7/Modified site: amidated carboxyl end (Leu) #status experimental

Query Match 35.1%; Score 13; DB 2; Length 7;
Best Local Similarity 66.7%; Pred. No. 2.2e+05;
Matches 2: Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 4 VPM 6
1 1 1
Db 2 MPM 4

RESULT 5

A28340
myomodulin - California sea hare
C:Species: Aplysia californica (California sea hare)
C:Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 18-Jun-1993
C:Accession: A28340
R:Cropper, E.C.; Tenenbaum, R.; Kolks, M.A.G.; Kupfermann, I.; Weiss, K.R.
Proc. Natl. Acad. Sci. U.S.A. 84, 5483-5486, 1987
A:Title: Myomodulin: A bioactive neuropeptide present in an identified cholinergic buccal
A:Reference number: A28340; MUID:87261010
A:Accession: A28340
A:Molecule type: protein

A:Residues: 1-7 <CRO>

Query Match 35.1%; Score 13; DB 2; Length 7;
Best Local Similarity 66.7%; Pred. No. 2.2e+05;
Matches 2: Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 5 PMN 7
1 1 1
Db 1 PMS 3

RESULT 6

A27696
contraction-inhibiting peptide I - blue mussel
C:Species: Mytilus edulis (blue mussel)
C:Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 08-Dec-1995
C:Accession: A27696
R:Hirata, T.; Kubota, I.; Iwasawa, N.; Takabatake, I.; Ikeda, T.; Muneoka, Y.
Biochem. Biophys. Res. Commun. 152, 1376-1382, 1988
A:Title: Structures and actions of Mytilus inhibitory peptides.
A:Reference number: A90142; MUID:88240357
A:Accession: A27696
A:Molecule type: protein
A:Residues: 1-6 <HIR>
C:Keywords: amidated carboxyl end
F:6/Modified site: amidated carboxyl end (Val) #status experimental

Query Match 32.4%; Score 12; DB 2; Length 6;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 2: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 PM 6
1 1 1
Db 3 PM 4

RESULT 7

B27696
contraction-inhibiting peptide II - blue mussel
C:Species: Mytilus edulis (blue mussel)
C:Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 08-Dec-1995
C:Accession: B27696
R:Hirata, T.; Kubota, I.; Iwasawa, N.; Takabatake, I.; Ikeda, T.; Muneoka, Y.
Biochem. Biophys. Res. Commun. 152, 1376-1382, 1988
A:Title: Structures and actions of Mytilus inhibitory peptides.
A:Reference number: A90142; MUID:88240357
A:Accession: B27696
A:Molecule type: protein
A:Residues: 1-6 <HIR>
C:Keywords: amidated carboxyl end
F:6/Modified site: amidated carboxyl end (Val) #status experimental

Query Match 32.4%; Score 12; DB 2; Length 6;
Best Local Similarity 100.0%; Pred. No. 2.2e+05;
Matches 2: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 PM 6
1 1 1
Db 3 PM 4

RESULT 8

A61300
22K superhelical DNA-binding protein - Escherichia coli (fragment)
C:Species: Escherichia coli
C:Date: 17-Jul-1994 #sequence_revision 17-Jul-1994 #text_change 07-May-1999
C:Accession: A61300
R:Kishi, F.; Edina, Y.; Miki, T.; Nakazawa, T.; Nakazawa, A.
J. Biochem. 92, 1059-1068, 1982
A:Title: Purification and characterization of a protein from Escherichia coli which f

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 6, 2002, 11:07:52 ; Search time 79.01 Seconds
(without alignments)
68.795 Million cell updates/sec

Title: US-09-897-042-13

Perfect score: 24

Sequence: 1 atgaactgtaccgatgaactaa 24

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues 281400

Total number of hits satisfying chosen parameters:
Minimum DB seq length: 0
Maximum DB seq length: 24

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database :

1: /cgn2_6/prodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/prodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/prodata/2/ina/5A_COMB.seq:*
4: /cgn2_6/prodata/2/ina/5B_COMB.seq:*
5: /cgn2_6/prodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/prodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	13.6	56.7	20	4	US-09-177-249-316
2	12.4	51.7	18	3	US-08-438-506-5
3	12.4	51.7	20	1	US-08-325-071-52
4	12.4	51.7	20	4	US-08-461-004A-52
5	12.4	51.7	21	1	US-08-325-071-54
6	12.4	51.7	21	4	US-08-461-004A-54
7	12.4	51.7	24	1	US-07-783-861C-13
8	12	50.0	24	2	US-08-805-918-19
9	12	50.0	24	2	US-08-805-918-19
10	12	50.0	24	2	US-08-805-918-19
11	11.6	48.3	19	1	US-08-811-028-15
12	11.6	48.3	19	2	US-08-110-294A-44
13	11.6	48.3	19	2	US-08-389-926-44
14	11.6	48.3	20	2	US-08-117-952-82
15	11.6	48.3	21	1	US-08-075-533-15
16	11.6	48.3	21	1	US-08-125-012-18
17	11.6	48.3	21	1	US-08-221-579A-12
18	11.6	48.3	21	1	US-08-783-818-18
19	11.6	48.3	21	2	US-08-453-349-18
20	11.6	48.3	21	2	US-08-948-176-15
21	11.6	48.3	21	2	US-08-704-701-12
22	11.6	48.3	21	2	US-08-979-385B-16
23	11.6	48.3	21	4	US-09-321-461-12
24	11.6	48.3	21	5	PCT-US91-09160-15
25	11.6	48.3	24	1	US-08-316-293-65
26	11.6	48.3	24	1	US-08-316-293-66
27	11.4	47.5	16	1	US-08-325-071-51

C 28	11.4	47.5	16	4	US-08-461-004A-51	Sequence 51, Appl
C 29	11.4	47.5	18	3	US-08-849-075-6	Sequence 6, Appl
C 30	11.4	47.5	19	3	US-09-175-219A-1	Sequence 1, Appl
C 31	11.4	47.5	20	1	US-08-096-182A-13	Sequence 13, Appl
C 32	11.4	47.5	20	1	US-08-877-109-13	Sequence 13, Appl
C 33	11.4	47.5	20	2	US-08-734-941-13	Sequence 3, Appl
C 34	11.4	47.5	20	3	US-08-798-760-13	Sequence 13, Appl
C 35	11.4	47.5	20	5	PCT-US94-08327-13	Sequence 13, Appl
C 36	11.4	47.5	21	1	US-08-586-274-4	Sequence 4, Appl
C 37	11.4	47.5	21	1	US-08-586-274-24	Sequence 24, Appl
C 38	11.4	47.5	21	3	US-09-116-294-15	Sequence 15, Appl
C 39	11.4	47.5	21	3	US-08-849-075-2	Sequence 2, Appl
C 40	11.4	47.5	21	4	US-09-537-357-50	Sequence 50, Appl
C 41	11.4	47.5	22	4	US-08-949-155-15	Sequence 15, Appl
C 42	11.4	47.5	24	1	US-07-783-861C-12	Sequence 12, Appl
C 43	11.4	47.5	24	6	5256545-41	Sequence 4, Appl
C 44	11.2	46.7	17	4	US-09-068-860-4	Sequence 12, Appl
C 45	11.2	46.7	17	4	US-09-068-860-12	Sequence 12, Appl
C 46	11.2	46.7	21	1	US-08-075-533-16	Sequence 16, Appl
C 47	11.2	46.7	21	4	US-08-948-176-16	Sequence 16, Appl
C 48	11.2	46.7	21	4	US-09-050-159-60	Sequence 60, Appl
C 49	11.2	46.7	21	5	PCT-US91-09160-16	Sequence 16, Appl
C 50	11.2	46.7	22	1	US-08-717-526-21	Sequence 21, Appl

ALIGNMENTS

RESULT 1
US-09-177-249-316/C
; Sequence 316, Application US/09177249
; Patent No. 6229064
; GENERAL INFORMATION:
; APPLICANT: Fischer, Robert L.
; APPLICANT: Fischer, Robert L.
; APPLICANT: Ohad, Nir
; APPLICANT: Kiyosue, Tomohiro
; APPLICANT: Yadegari, Ramon
; APPLICANT: Margossian, Linda
; APPLICANT: Harada, John
; APPLICANT: Goldberg, Robert B.
; TITLE OF INVENTION: Nucleic Acids That Control Seed and Fruit
; FILE OF INVENTION: Development In Plants
; FILE REFERENCE: 023070-086120US
; CURRENT APPLICATION NUMBER: US/09/177,249
; CURRENT FILING DATE: 1998-10-22
; EARLIER APPLICATION NUMBER: US 09/071,838
; NUMBER OF SEQ ID NOS: 324
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 316
; LENGTH: 20
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: primer
US-09-177-249-316

Query Match 56.7% ; Score 13.6 ; DB 4 ; Length 20 ;
Best local Similarity 80.0% ; Pred. No. 5.1e+02 ;
Matches 16 ; Conservative 0 ; Mismatches 4 ; Indels 0 ; Gaps 0 ;

QY 3 gaactgtaccgatgaact 22
Db 20 GAACCTCATAGCGATGACT 1

RESULT 2
US-08-438-506-5
; Sequence 5, Application US/08438506
; Patent No. 6001562

```

: GENERAL INFORMATION:
: APPLICANT: Milosavljevic, Aleksandar
: TITLE OF INVENTION: DNA SEQUENCE SIMILARITY RECOGNITION BY
: TITLE OF INVENTION: HYBRIDIZATION TO SHORT OLIGOMERS
: NUMBER OF SEQUENCES: 8
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Arnold, White & Durkee
: STREET: P.O. Box 4433
: CITY: Houston
: STATE: Texas
: COUNTRY: United States of America
: ZIP: 77210
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/438,506
: FILING DATE: Concurrently Herewith
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Highlander, Steven L.
: REGISTRATION NUMBER: 37,642
: REFERENCE/DOCKET NUMBER: ARCD:153/HYL
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (512) 418-3000
: TELEFAX: (713) 789-2679
: TELEX: 79-0924
: INFORMATION FOR SEQ ID NO: 5:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 18 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: US-08-438-506-5

Query Match          51.7%; Score 12.4; DB 3; Length 18;
Best Local Similarity 92.9%; Pred. No. 1.9e+03;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 11 taccgatgaactaa 24
Db 4 TACCGATGAGCTAA 17

RESULT 3
US-08-325-071-52
: Sequence 52, Application US/08325071
: Patent No. 5587311
: GENERAL INFORMATION:
: APPLICANT: COBON, Stewart Gary
: APPLICANT: MOORE, Joanna Terry
: APPLICANT: JOHNSON, Law Anthony York
: APPLICANT: WILADSEN, Peter
: APPLICANT: KEMP, David Harold
: APPLICANT: SRISKANTHA, Alagacone
: APPLICANT: RIDING, George Alfred
: APPLICANT: RAND, Keith No. 5587311man
: TITLE OF INVENTION: DNA Encoding A Cell Membrane
: TITLE OF INVENTION: Glycoprotein Of A Tick Gut
: NUMBER OF SEQUENCES: 71
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Foley & Lardner
: STREET: 3000 K Street, N.W.
: CITY: Washington, D.C.
: COUNTRY: USA
: ZIP: 20007-5109
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
```

```

: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/325,071
: FILING DATE: 14-OCT-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/062,109
: FILING DATE: 17-MAY-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/926,368
: FILING DATE: 07-AUG-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 07/242,196
: FILING DATE: 06-JUL-1988
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/AU87/00401
: FILING DATE: 27-NOV-1987
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: AU P14912
: FILING DATE: 16-OCT-1987
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: AU P12570
: FILING DATE: 19-JUN-1987
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: AU P9196
: FILING DATE: 27-NOV-1986
: ATTORNEY/AGENT INFORMATION:
: NAME: BENT, Stephen A.
: REGISTRATION NUMBER: 29,768
: REFERENCE/DOCKET NUMBER: 60042/111 BIAU
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 202 672 5300
: TELEFAX: 202 672 5399
: TELEX: 904136
: INFORMATION FOR SEQ ID NO: 52:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 20 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: ANTI-SENSE: YES
: US-08-325-071-52

Query Match          51.7%; Score 12.4; DB 1; Length 20;
Best Local Similarity 92.9%; Pred. No. 1.9e+03;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 gaacttggtaccga 16
Db 4 GAAC TGCGGTACCGA 17

RESULT 4
US-08-461-004A-52
: Sequence 52, Application US/08461004A
: Patent No. 6235283
: GENERAL INFORMATION:
: APPLICANT: COBON, Stewart Gary
: APPLICANT: MOORE, Joanna Terry
: APPLICANT: JOHNSON, Law Anthony York
: APPLICANT: WILADSEN, Peter
: APPLICANT: KEMP, David Harold
: APPLICANT: SRISKANTHA, Alagacone
: APPLICANT: RIDING, George Alfred
: APPLICANT: RAND, Keith No. 6235283man
: TITLE OF INVENTION: DNA Encoding A Cell Membrane
: TITLE OF INVENTION: Glycoprotein Of A Tick Gut
: NUMBER OF SEQUENCES: 71
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Foley & Lardner
: STREET: 3000 K Street, N.W.
: CITY: Washington, D.C.
: COUNTRY: USA
```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 6, 2002, 12:24:52 ; Search time 142.39 Seconds
(without alignments)
144.503 Million cell updates/sec

Title: US-09-897-042-13
Perfect score: 24
Sequence: 1 atgaactgtgacgatgaactaa 24

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues
Total number of hits satisfying chosen parameters: 607738

Minimum DB seq length: 0
Maximum DB seq length: 24

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : N_Geneseq_1101.*
1: /SID2/gcgdata/geneseq/NA1980.DAT.*
2: /SID2/gcgdata/geneseq/NA1981.DAT.*
3: /SID2/gcgdata/geneseq/NA1982.DAT.*
4: /SID2/gcgdata/geneseq/NA1983.DAT.*
5: /SID2/gcgdata/geneseq/NA1984.DAT.*
6: /SID2/gcgdata/geneseq/NA1985.DAT.*
7: /SID2/gcgdata/geneseq/NA1986.DAT.*
8: /SID2/gcgdata/geneseq/NA1987.DAT.*
9: /SID2/gcgdata/geneseq/NA1988.DAT.*
10: /SID2/gcgdata/geneseq/NA1989.DAT.*
11: /SID2/gcgdata/geneseq/NA1990.DAT.*
12: /SID2/gcgdata/geneseq/NA1991.DAT.*
13: /SID2/gcgdata/geneseq/NA1992.DAT.*
14: /SID2/gcgdata/geneseq/NA1993.DAT.*
15: /SID2/gcgdata/geneseq/NA1994.DAT.*
16: /SID2/gcgdata/geneseq/NA1995.DAT.*
17: /SID2/gcgdata/geneseq/NA1996.DAT.*
18: /SID2/gcgdata/geneseq/NA1997.DAT.*
19: /SID2/gcgdata/geneseq/NA1998.DAT.*
20: /SID2/gcgdata/geneseq/NA1999.DAT.*
21: /SID2/gcgdata/geneseq/NA2000.DAT.*
22: /SID2/gcgdata/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	13	54.2	20	18	AAT60152
2	12.8	53.3	24	19	AAV34719
3	12.6	52.5	20	21	AAV58055
4	12.6	52.5	22	20	AAZ40501
5	12.6	52.5	22	20	AAZ08927
6	12.4	51.7	18	21	AAZ46447
7	12.2	50.8	17	20	AAV17376
8	12.2	50.8	18	22	AAZ9290
9	12.2	50.8	20	18	AAT89747
10	12.2	50.8	20	20	AAZ96327
11	12	50.0	20	22	AAZ5655

12	12	50.0	21	12	AAQ15384	Toxicogenic Clostrid
13	12	50.0	21	19	AAW64140	Rat activin recept
14	11.8	49.2	17	20	AAZ32175	Probe specific for
15	11.8	49.2	19	21	AAH83059	cdk6 ribozyme bind
16	11.8	49.2	19	22	AAH58221	Cell-cycle depende
17	11.8	49.2	20	20	AAZ33833	PCR primer used to
18	11.8	49.2	21	22	AAH62391	Voltage gated Ca c
19	11.8	49.2	21	22	AAH63472	Oligonucleotide CL
20	11.6	48.3	19	18	AAT77582	Wheat microsatelli
21	11.6	48.3	20	16	AAO82082	Chromosome 11 (loc
22	11.6	48.3	21	13	AAO26534	Acyl-ACP thioester
23	11.6	48.3	21	15	AAO67153	Primer for amplify
24	11.6	48.3	21	16	AAT00841	HIV gag/pol gene f
25	11.6	48.3	21	17	AAT32899	HIV-1 HXB2 gag/pol
26	11.6	48.3	21	18	AAT77580	Wheat microsatelli
27	11.6	48.3	21	20	AAH80893	Acyl-ACP thioester
28	11.6	48.3	21	21	AAH94143	Adenovirus 5 E2/E3
29	11.6	48.3	21	22	AAZ6404	Human gene single
30	11.6	48.3	22	21	AAH96903	RACE PCR primer us
31	11.6	48.3	22	22	AAZ03611	PCR primer rend9b1
32	11.6	48.3	23	20	AAZ6142	Primer for generat
33	11.6	48.3	24	13	AAO23101	HPV11 probe ICR2B
34	11.6	48.3	24	13	AAO23102	HPV11 probe ICR2B
35	11.6	48.3	24	17	AAT29880	Human papillomavir
36	11.6	48.3	24	17	AAT29891	Human papillomavir
37	11.4	47.5	19	21	AAZ73913	Aryl hydrocarbon m
38	11.4	47.5	19	20	AAZ73913	Human biallelic ma
39	11.4	47.5	20	16	AAO85297	5' oligo primer fo
40	11.4	47.5	20	18	AAT60151	Primer #4 for tyro
41	11.4	47.5	20	18	AAT60154	Primer #7 for tyro
42	11.4	47.5	20	18	AAT60153	Primer #6 for tyro
43	11.4	47.5	20	20	AAZ01097	Iga-binding fragme
44	11.4	47.5	21	17	AAZ9325	Methicillin resist
45	11.4	47.5	22	19	AAZ29168	Nucleotide sequenc
46	11.4	47.5	24	21	AAZ87485	Human ADH7 gene ex
47	11.4	47.5	24	22	AAH46732	Type 11 phosphodi
48	11.2	46.7	17	20	AAV1940	Primer used for se
49	11.2	46.7	17	21	AAH58620	Primer used for pio
50	11.2	46.7	18	13	AAO22881	HCV-HC59 primer #1

ALIGNMENTS

RESULT 1	
1	AAT60152 standard; DNA: 20 BP.
ID	AAT60152
XX	25-NOV-1997 (first entry)
AC	Primer #5 for tyrosine 3-hydroxylase gene from vitelline cell cDNA.
XX	AT60152:
DT	XX
XX	XX
DE	XX
XX	XX
XX	XX
OS	Synthetic.
XX	XX
PN	W0971191-A1.
XX	27-MAR-1997.
PD	XX
XX	XX
XX	20-SEP-1996; 96WO-US15083.
PF	XX
XX	XX
XX	21-SEP-1995; 95US-0004115.
PR	XX
XX	XX
PA	(MILLER) MILLER I.
XX	XX
PI	Miller I;

DR WPI: 1997-202898/18.
 XX Secretion of therapeutic proteins - from transgenic schistosomes
 PT Into the blood of a animal host
 PS
 XX Disclosure: Page 37; 48pp; English.
 XX
 CC The sequences given in AAT60148-54 are primers which were used in the
 CC amplification of the schistosome tyrosine 3-hydroxylase (TY3H) from
 CC vitelline cell cDNA. The upstream primers (AAT60148-50) are based on
 CC drosophila TY3H amino acids 368-373, and the downstream primers
 CC (AAT60151-54) are based on drosophila TY3H amino acids 428-433. The
 CC amplified sequences may be used in the method of the invention for the
 CC production of transgenic schistosomes for secretion of particular gene
 CC products. The method comprises cloning the protein coding portion of a
 CC cDNA corresponding to the gene encoding the protein of interest into a
 CC plasmid having upstream and downstream promoter/enhancer sequences, and
 CC containing 5' and 3' untranslated regions of a schistosome gene that is
 CC expressed in the integument, vitelline or Mehli's glands, or ootype of
 CC schistosomes. The transgene DNA is microinjected into the pronuclei or
 CC cytoplasm of the zygotes of stage I schistosome eggs and the eggs are
 CC cultured to maturity in vitro. Miracidia are hatched and snails are
 CC infected with one miracidium each. The snails are cultured until
 CC cercariae of schistosome clones are produced. Transgenic clones are
 CC identified at the cercarial stage and maintained by sporocyst transfer.
 CC Laboratory animals are infected with one cercarium each. Transgenic
 CC schistosome clones that secrete high levels of the desired protein
 CC into the peripheral blood of the host are then identified. The method
 CC is used to produce therapeutic proteins in the blood stream of animals.
 CC Typical proteins to be produced include insulin, leptin, calcitonin,
 CC alpha 1 anti-trypsin, factor VIII, HIV co-receptor ligands and
 CC cholesterol ester transfer protein inhibitor. Use of schistosomes as
 CC intermediate vectors for gene therapy facilitates mass production,
 CC quality control, termination of treatment and dose titration. It avoids
 CC problems, such as low efficiency, possible infection by contaminating
 CC replication competent virus, potential recombination with host DNA,
 CC possible malignant transformation and need to individualise treatments,
 CC associated with use of viral vectors.
 XX
 S0 Sequence 20 BP; 3 A; 2 C; 5 G; 5 T; 5 other:
 / 2 tgaacttggtaccgatgaac 21
 || : || : || : || : || : ||
 Db 1 tgrtcygtgrtangsytgaac 20
 / 2
 || : || : || : || : || : ||
 Db 1 tgrtcygtgrtangsytgaac 20
 RESULT 2
 ID AAV34719/C
 XX AAV34719 standard; cDNA: 24 BP.
 XX
 XX AAV34719:
 XX
 DT 27-AUG-1998 (first entry)
 XX
 XX Human Fanconl I primer SP2.
 KW Fanconl; anaemia; disease; diagnosis; disorder; predisposition; tumour;
 KW cell cycle; cell activation; DNA repair; cytopaenia; treatment; effector;
 KW prevention; gene therapy; proliferative disease; therapeutic agent;
 KW primer; ss.
 XX
 XX Synthetic.
 OS
 OS Homo sapiens.
 XX
 XX W09816637-A1.
 XX
 XD 23-APR-1998.

```

XX 08-OCT-1997; 97WO-EP05543.
PF
XX
XX 11-OCT-1996; 96EP-0116336.
PR
XX
XX (BOEF ) BOEHRINGER MANNHEIM GMBH.
PA
XX
XX Kubbles M, Machl A, Planitzer S;
PI
XX
XX WPI; 1998-261040/23.
DR
XX
XX Nucleic acid corresponding to Fanconi anaemia gene - used for
PT diagnosis, treatment and prevention of diseases involving disordered
PT cell cycle progression, cytopaenia etc.
PT
XX
XX Example 4; Page 15; 37pp; German.
PS
XX
XX AAV34718-734722 are primers used in the isolation of a gene associated
CC with Fanconi anaemia. This sequence can be used for diagnosing disease,
CC or predisposition to it, associated with disorders of cell cycling,
CC activation or cell-cycle progression; DNA repair; cytopaenia; or
CC formation/progression of tumours, or to treat or prevent such
CC conditions, particularly by gene therapy, especially of cytopaenia,
CC tumours or other proliferative diseases. Cells that express this protein
CC can be used to identify specific effectors (potentially useful as
CC therapeutic agents).
CC
XX
XX Sequence 24 BP; 8 A; 6 C; 5 G; 5 T; 0 other;
SQ
XX
XX Query Match 53.3%; Score 12.8; DB 19; Length 24;
XX Best Local Similarity 87.5%; Pred. No. 3.6e+03;
XX Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
XX 7 ttgtgacgatgaact 22
QY ||||||| ||||| ||
XX Db 24 TTGGTACAGATGATCT 9
XX
XX RESULT 3
XX ID AAA58055/c
XX AAA58055 standard; DNA; 20 BP.
XX
XX AAA58055;
XX
XX 10-OCT-2000 (first entry)
DX
XX
XX Human interleukin-15 (IL-15) antisense oligonucleotide, ODN 10.
XX
XX Human interleukin-15; IL-15; antisense oligonucleotide;
XX expression inhibition; T-cell mediated immune response;
XX autoimmune disorder; inflammatory polyarthropathy; rheumatoid arthritis;
XX transplant rejection; graft versus host disease; lupus erythematosus;
XX asthma; inflammatory bowel disease; gene therapy; immunosuppressive;
XX antiinflammatory; 3' UTR; 3' untranslated region; ss.
XX
XX Homo sapiens.
XX
XX WO200028019-A2.
XX
XX 18-MAY-2000.
XX
XX 04-NOV-1999; 99WO-IL00589.
XX
XX 05-NOV-1998; 98IL-0126919.
XX
XX (UYNE ) UNIV BEN-GURION NEGEV RES & DEV.
XX (MORR-) MOR RES APPL LTD.
XX
XX Doudevant A, Chaimovitz C;
XX
XX WPI; 2000-376536/32.
XX

```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 6, 2002, 11:01:22 ; Search time 1365.98 Seconds
(without alignments)
188.801 Million cell updates/sec

Title: US-09-897-042-13
Perfect score: 24
Sequence: 1 atgaacttgctacgatgaactaa 24

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues
Total number of hits satisfying chosen parameters: 13210

Minimum DB seq length: 0
Maximum DB seq length: 24

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database :

EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estlin:*
4: em_estlom:*
5: em_estpl:*
6: em_estlda:*
7: em_estlro:*
8: em_estlov:*
9: em_hlc:*
10: gb_estl:*
11: gb_est2:*
12: gb_hlc:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rpd:*
20: em_gss_vit:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	11.2	46.7	24	AZ581676	1M0370N06
2	10.8	45.0	19	AZ340577	1M0072H21
3	10.8	45.0	21	AZ663981	1M0543J22
4	10.2	42.5	22	AZ501093	1M0339L13
5	10.2	42.5	24	AZ360525	1M0103N13
6	10.2	41.7	19	AU061154	1M0103N13
7	10	41.7	20	TA339H110	1M0103N13
8	9.8	40.8	20	AZ620235	1M0452E23
9	9.8	40.8	22	AZ779213	1M0051L18
10	9.6	40.0	21	AZ770212	1M0571G10
11	9.6	40.0	24	TA116E02P	1M0571G10
12	9.4	39.2	19	AZ836789	2M0131I18

C	13	9.4	39.2	21	13	AZ828967	AZ828967 2M0106N15
C	14	9.4	39.2	22	13	TA20G06P	AL453552 T. brucei
C	15	9.4	39.2	22	13	TA114G120	AL462979 T. brucei
C	16	9.4	39.2	23	13	AZ333221	AZ333221 1M0062O11
C	17	9.4	39.2	23	13	AZ459779	AZ459779 1M0264A22
C	18	9.2	38.3	17	2	HSMD007757	AL042907 Homo sapi
C	19	9.2	38.3	19	13	AZ787717	AZ787717 2M0034A21
C	20	9.2	38.3	21	13	AZ609424	AZ609424 1M0434O16
C	21	9.2	38.3	21	13	AZ828967	AZ828967 2M0106N15
C	22	9.2	38.3	22	13	AZ394329	AZ394329 1M0157B21
C	23	9	37.5	19	13	AZ399413	AZ399413 1M0165E16
C	24	9	37.5	19	13	AZ815827	AZ815827 2M0084K23
C	25	9	37.5	19	13	AZ991317	AZ991317 2M0275K24
C	26	9	37.5	20	13	AZ658617	AZ658617 1M0535E04
C	27	9	37.5	21	13	AZ768984	AZ768984 1M0569C16
C	28	9	37.5	22	13	AZ635609	AZ635609 1M0493C08
C	29	9	37.5	22	13	AZ840373	AZ840373 2M0136G21
C	30	9	37.5	22	13	AZ871339	AZ871339 2M0184G09
C	31	9	37.5	23	13	AZ308507	AZ308507 1M0011L02
C	32	9	37.5	24	13	TA294D030	AL486340 T. brucei
C	33	8.8	36.7	19	13	AZ831033	AZ831033 2M0110O01
C	34	8.8	36.7	22	13	AZ429443	AZ429443 1M0213G21
C	35	8.8	36.7	22	13	AZ450734	AZ450734 1M0249B10
C	36	8.8	36.7	22	13	AZ598849	AZ598849 1M0399A09
C	37	8.8	36.7	22	13	TA19D050	AL453318 T. brucei
C	38	8.8	36.7	23	13	AZ370804	AZ370804 1M0121B20
C	39	8.8	36.7	23	13	AZ828383	AZ828383 2M0105O15
C	40	8.8	36.7	24	13	AZ313141	AZ313141 1M0029D19
C	41	8.8	36.7	24	13	AZ342545	AZ342545 1M0075I11
C	42	8.8	36.7	24	13	AZ804947	AZ804947 2M0068B07
C	43	8.6	35.8	20	13	AZ936985	AZ936985 2M0195E06
C	44	8.6	35.8	20	13	AZ943793	AZ943793 2M0204A07
C	45	8.6	35.8	21	13	AZ453408	AZ453408 1M0254H12
C	46	8.6	35.8	22	13	AZ394329	AZ394329 1M0157B21
C	47	8.6	35.8	23	13	AZ805705	AZ805705 2M0067D18
C	48	8.6	35.8	23	13	TA195D060	AL476956 T. brucei
C	49	8.6	35.8	24	13	AZ628737	AZ628737 1M0481E09
C	50	8.4	35.0	19	13	AZ312945	AZ312945 1M0029P03

ALIGNMENTS

RESULT	1	AZ581676	24 bp	DNA	GSS	13-DEC-2000
LOCUS	1M0370N06R	Mouse 10kb plasmid UUGC1M library	Mus musculus genomic			
DEFINITION	clone UUGC1M0370N06 R, DNA sequence.					
ACCESSION	AZ581676					
VERSION	AZ581676.1	GI:11696926				
KEYWORDS	GSS.					
SOURCE	house mouse.					
ORGANISM	Mus musculus					
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.					
AUTHORS	1 (bases 1 to 24) Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmood,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D., Weiss,R. Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts Unpublished (2000) Contact: Robert B. Weiss University of Utah Genome Center Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLG, UT 84112, USA Tel.: 801 585 5606 Fax: 801 585 7177 Email: ddunn@genetics.utah.edu Insert Length: 10000 Std Error: 0.00 Plate: 0370 row: N column: 06					
TITLE	JOURNAL					
COMMENT						

```

Seq primer: CACACAGGAAACAGCTATGACCC
Class: Plasmid ends
High quality sequence stop: 24.
      location/Qualifiers
1..24
FEATURES
      source

```

BASE COUNT
ORIGIN

5 a 4 c 4 g 11 t

```

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/cclone="UUGC1M0370N06"
/cclone_11b="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/notes="Vector: pWD42ny; Purified genomic DNA from M.
musculus C57BL/6J (male); was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptor DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pWD42 (g11473211419b1F129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptor mouse DNA was annealed to
adaptor vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

```

	Query Match	46.7%	Score 11.2:	DB 13:	Length 24:
	Best Local Similarity	81.2%:	Pred. No. 3.	8e+05:	
	Matches 13:	Conservative	0:	Mismatches 3:	Indels 0:
Gy	9 ggtacgcgatgaactaa	24			
Db	18 GCTATCGATGAATTAA	3			

RESULT	2
LOCUS	AZ340577
DEFINITION	AZ340577 19 bp DNA GSS 29-SEP-2000
ACCESSION	U00072H21R Mouse 10kb plasmid UGCC1M library Mus musculus genomic
VERSION	AZ340577
KEYWORDS	AZ340577.1 GI:10415969
SOURCE	GSS.
ORGANISM	house mouse. Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclirognathi; Muridae; Murinae; Mus 1 (bases 1 to 19)
AUTHORS	Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly, M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D., Weiss,R.
TITLE	Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL	Unpublished (2000)
COMMENT	Contact: Robert B. Weiss

JOURNAL COMMENT

Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 508, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00

```

Plate: 0072  row: H  column: 21
Seq primer: CACACAGGAAACACCTATGACC
Class: Plasmid ends
High quality sequence stop: 19.
Location/Qualifiers
1..19
FEATURES
SOURCE
```

```

BASE COUNT
ORIGIN
5 a      4 c      6 g      4 t
      /organism="Mus musculus"
      /strain="C57BL/6J"
      /db_xref="taxon:10090"
      /clone="UUGC1M0072H21"
      /clone_1b="Mouse 10kb plasmid UUGC1M library"
      /sex="Male"
      /lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
      /note="Vector: pMD42ny: Purified genomic DNA from M.
      musculus C57BL/6J (male) was obtained from the Jackson
      Laboratory Mouse DNA Resource
      (http://www.jax.org/resources/documents/dnares/). The DNA
      was hydrodynamically sheared by repeated passage through a
      0.005 inch orifice at constant velocity. The sheared DNA
      was blunt end-repaired with T4 DNA polymerase and T4
      polynucleotide kinase. Adaptor oligonucleotides were
      ligated to the blunt ends in high molar excess. The
      adaptor DNA was purified and size-selected for a 9.5 to
      10.5 kb range using preparative agarose gel
      electrophoresis. Vector DNA was prepared from a derivative
      of pMD42 (g11473211419b1Afr29072.1), a copy-number
      inducible derivative of plasmid R1. The vector was ligated
      with adaptors complementary to the insert adaptors and
      purified. The sheared, adaptor mouse DNA was annealed to
      adaptor vector DNA, and transformed into
      chemically-competent E. coli XL10-Gold (Stratagene) cells
      and selected for ampicillin resistance."

```

Query Match	45.0%;	Score 10.8;	DB 13;	Length 19;
Best Local Similarity	85.7%;	Pred. No. 5.7e+05;		
Matches 12; Conservative	0;	Mismatches 2;	Indels 0;	Gaps 0;
QY	9	gtatccagatgaact	22	
	3	GGTAACGCTGAAC	16	

RESULT	3
AZ663981/c	
LOCUS	AZ663981 21 bp DNA 14-DEC-2000
DEFINITION	U05033J322R Mouse 10kb plasmid UGCG1M library Mus musculus genomic clone UGCG1M0543J22 R, plasma sequence.
ACCESSION	AZ663981
VERSION	AZ663981.1 GI:11801127
KEYWORDS	GSS.
SOURCE	house mouse.
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus 1 (bases 1 to 21)
AUTHORS	Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.
TITLE	Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL	Unpublished (2000)
COMMENT	Contact: Robert B. Weiss

JOURNAL COMMENT

Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 508, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel.: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 6, 2002, 11:02:32 ; Search time 1434.12 seconds
(without alignments)
276.080 Million cell updates/sec

Title: US-09-897-042-13

Perfect score: 24
Sequence: 1 atgaactgtgacgacgatgaactaa 24

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 309068

Minimum DB seq length: 0
Maximum DB seq length: 24

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 50 summaries

Database : GenBank:
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_cm:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_om:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vl:*
30: em_htgo_hum:*
31: em_htgo_inv:*
32: em_htgo_rnd:*
33: em_htg_hum:*
34: em_htg_inv:*
35: em_htg_rnd:*
36: em_htg_other:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	13.6	56.7	20	6	ARI50940 Sequence
C 2	12.6	52.5	20	6	AX024726 Sequence
C 3	12.4	51.7	18	6	AR093435 Sequence
C 4	12.4	51.7	20	6	ARI52501 Sequence
C 5	12.4	51.7	20	6	I32378 Sequence 52
C 6	12.4	51.7	21	6	ARI52503 Sequence
C 7	12.4	51.7	21	6	I32380 Sequence 54
C 8	12.4	51.7	24	6	I15196 Sequence 13
C 9	12.4	50.0	20	6	AX082987 Sequence
C 10	12.2	50.0	21	6	AX029243 Sequence
C 11	12.2	50.0	24	6	AR069541 Sequence
C 12	12.2	50.0	24	6	E16933 Linker DNA.
C 13	12.2	50.0	24	6	E16936 Linker DNA.
C 14	11.8	49.2	19	6	AX129427 Sequence
C 15	11.8	49.2	21	6	AX095881 Sequence
C 16	11.8	49.2	24	6	AX085594 Sequence
C 17	11.6	48.3	18	6	A35582 Synthetic h
C 18	11.6	48.3	19	6	AR034541 Sequence
C 19	11.6	48.3	19	6	AR048641 Sequence
C 20	11.6	48.3	20	6	AR066734 Sequence
C 21	11.6	48.3	21	6	AR001369 Sequence
C 22	11.6	48.3	21	6	AR0044856 Sequence
C 23	11.6	48.3	21	6	AR052242 Sequence
C 24	11.6	48.3	21	6	AR078349 Sequence
C 25	11.6	48.3	21	6	AR085200 Sequence
C 26	11.6	48.3	21	6	AR138120 Sequence
C 27	11.6	48.3	21	6	AX036805 Sequence
C 28	11.6	48.3	21	6	I33906 Sequence 18
C 29	11.6	48.3	21	6	I17054 Sequence 65
C 30	11.6	48.3	21	6	I17055 Sequence 66
C 31	11.6	48.3	21	6	ARI52500 Sequence
C 32	11.6	48.3	21	6	I32377 Sequence 51
C 33	11.6	48.3	21	6	ARI10519 Sequence
C 34	11.6	48.3	21	6	AR004782 Sequence
C 35	11.6	48.3	21	6	AR032321 Sequence
C 36	11.6	48.3	21	6	I13582 Sequence 13
C 37	11.6	48.3	21	6	AX096502 Sequence
C 38	11.6	48.3	21	6	AX154194 Sequence
C 39	11.6	48.3	21	6	E12099 PCR primer
C 40	11.6	48.3	21	6	I86678 Sequence 4
C 41	11.6	48.3	21	6	I86698 Sequence 24
C 42	11.6	48.3	21	6	AX067767 Sequence
C 43	11.6	48.3	21	6	I15195 Sequence 12
C 44	11.6	48.3	21	6	E15729 PCR primer
C 45	11.6	48.3	21	6	ARI37313 Sequence
C 46	11.6	48.3	21	6	AR037435 Sequence
C 47	11.6	48.3	21	6	AX095991 Sequence
C 48	11.6	48.3	21	6	A61184 Sequence 21
C 49	11.6	48.3	21	6	A76352 Sequence 18
C 50	11.6	48.3	21	6	AR031334 Sequence

ALIGNMENTS

RESULT 1
LOCUS ARI50940/c 20 bp DNA
DEFINITION Sequence 316 from patent US 6229064.
ACCESSION ARI50940
VERSION ARI50940.1 GI:15115531
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 20)
AUTHORS Fischer R.L., Ohad N., Kiyosue T., Yadegaril R., Margossian L.,
Harada J. and Goldberg R.B.
TITLE Nucleic acids that control endosperm development in plants
JOURNAL Patent: US 6229064-A 316 08-MAY-2001;

closed

15 not run -

too large

FEATURES Location/Qualifiers
source 1..20
/organism="unknown"
BASE COUNT 5 a 4 c 5 g 6 t
ORIGIN

Query Match 56.7%; Score 13.6; DB 6; Length 20;
Best Local Similarity 80.0%; Pred. No. 4e+04;
Matches 16; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 3 gaacttggtaccgatgaact 22
||||| || ||||| ||
Db 20 GAACCTCATGACGATGACCT 1

RESULT 2
AX024726/c 20 bp DNA PAT 15-SEP-2000
LOCUS AX024726
DEFINITION Sequence 12 from Patent WO0028019.
ACCESSION AX024726
VERSION AX024726.1 GI:10184805
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 20)
AUTHORS Doudevani, A. and Chalmovitz, C.
TITLE Antisense oligomer
JOURNAL MOR RESEARCH APPLIC LTD (IL) ; DOUDEVANI AMOS (IL) ; UNIV BEN
CURION (IL) ; CHALMOVITZ CIDIO (IL)

FEATURES Location/Qualifiers
source 1..20
/organism="Homo sapiens"
BASE COUNT 5 a 4 c 5 g 6 t
ORIGIN

Query Match 52.5%; Score 12.6; DB 6; Length 20;
Best Local Similarity 78.9%; Pred. No. 1.3e+05;
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 4 aacttggtaccgatgaact 22
||||| || ||||| ||
, 20 ACCTTGATCAGATGACT 2

RESULT 3
AR093435 18 bp DNA PAT 08-SEP-2000
LOCUS AR093435
DEFINITION Sequence 5 from patent US 6001562.
ACCESSION AR093435
VERSION AR093435.1 GI:10020184
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 18)
AUTHORS Milosavljevic, A.
TITLE DNA sequence similarity recognition by hybridization to short
oligomers
JOURNAL Patent: US 6001562-A 5 14-DEC-1999;
FEATURES Location/Qualifiers
source 1..18
/organism="unknown"

BASE COUNT 6 a 3 c 4 g 5 t
ORIGIN

Query Match 51.7%; Score 12.4; DB 6; Length 18;

Best Local Similarity 92.9%; Pred. No. 1.6e+05;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 11 tacgatgaactaa 24
||||| ||||| |||||
Db 4 TACGATGAGCTAA 17

RESULT 4
AR152501 20 bp DNA PAT 08-AUG-2001
LOCUS AR152501
DEFINITION Sequence 52 from patent US 6235283.
ACCESSION AR152501
VERSION AR152501.1 GI:15120033
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 20)
AUTHORS Cobon, G., Stewart, Moore, J., Terry, Johnston, L., Anthony, York, Willadsen, P., Kemp, D., Harold, Sriskantha, A., Riding, G., Alfred and Rand, K. Norman.
TITLE DNA encoding a cell membrane glycoprotein of a tick gut
JOURNAL Patent: US 6235283-A 52 22-MAY-2001;
FEATURES Location/Qualifiers
source 1..20
/organism="unknown"

BASE COUNT 6 a 5 c 6 g 3 t
ORIGIN

Query Match 51.7%; Score 12.4; DB 6; Length 20;
Best Local Similarity 92.9%; Pred. No. 1.6e+05;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 gaacttggtaccga 16
||||| ||||| |||||
Db 4 GAACRGGCTACCGA 17

RESULT 5
I32378 20 bp DNA PAT 06-FEB-1997
LOCUS I32378
DEFINITION Sequence 52 from patent US 5587311.
ACCESSION I32378
VERSION I32378.1 GI:1823169
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.

REFERENCE 1 (bases 1 to 20)
AUTHORS Cobon, G.S., Moore, J.T., Johnston, L.A.Y., Willadsen, P., Kemp, D.H., Sriskantha, A., Riding, G.A. and Rand, K.N.
TITLE DNA encoding a cell membrane glycoprotein of a tick gut
JOURNAL Patent: US 5587311-A 52 24-DEC-1996;
FEATURES Location/Qualifiers
source 1..20
/organism="unknown"

BASE COUNT 6 a 5 c 6 g 3 t
ORIGIN

Query Match 51.7%; Score 12.4; DB 6; Length 20;
Best Local Similarity 92.9%; Pred. No. 1.6e+05;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 gaacttggtaccga 16
||||| ||||| |||||
Db 4 GAACRGGCTACCGA 17

RESULT 6
AR152503

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 6, 2002, 09:15:15 ; Search time 71.9 seconds
(without alignments)
5.008 Million cell updates/sec

Title: US-09-897-042-18

Perfect score: 85

Sequence: 1 MNLVPMNPLVNMFMN 16

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database: Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCTUS.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	42	49.4	775	1 US-07-603-133B-13	Sequence 13, Appl
2	40	47.1	775	1 US-07-603-133B-14	Sequence 14, Appl
3	40	47.1	776	1 US-07-603-133B-17	Sequence 17, Appl
4	40	47.1	776	1 US-07-603-133B-20	Sequence 20, Appl
5	40	47.1	776	3 US-08-089-397A-15	Sequence 15, Appl
6	40	47.1	908	2 US-08-249-380-2	Sequence 2, Appl
7	38	44.7	510	1 US-08-249-112-1	Sequence 1, Appl
8	38	44.7	510	1 US-08-063-952-9	Sequence 9, Appl
9	38	44.7	694	4 US-08-701-240-2	Sequence 2, Appl
10	38	44.7	694	4 US-09-138-236-2	Sequence 2, Appl
11	38	44.7	775	1 US-07-603-133B-12	Sequence 12, Appl
12	38	44.7	776	1 US-07-603-133B-18	Sequence 18, Appl
13	38	44.7	2182	2 US-08-487-826B-16	Sequence 16, Appl
14	37	43.5	195	1 US-08-063-952-9	Sequence 9, Appl
15	37	43.5	195	5 PCT-US93-05704-9	Sequence 9, Appl
16	37	43.5	695	2 US-08-701-240-4	Sequence 4, Appl
17	37	43.5	695	4 US-09-138-236-4	Sequence 4, Appl
18	37	43.5	1375	4 US-09-210-361-4	Sequence 4, Appl
19	37	43.5	1475	3 US-09-007-999-2	Sequence 2, Appl
20	37	43.5	1475	4 US-09-210-361-2	Sequence 2, Appl
21	36	42.4	409	2 US-09-031-485-15	Sequence 15, Appl
22	36	42.4	409	2 US-08-847-429A-15	Sequence 15, Appl
23	36	42.4	409	3 US-09-075-215A-17	Sequence 17, Appl
24	36	42.4	409	3 US-09-065-474-15	Sequence 15, Appl
25	36	42.4	712	2 US-08-468-576B-17	Sequence 17, Appl
26	36	42.4	712	2 US-08-468-576B-17	Sequence 17, Appl
27	36	42.4	712	3 US-08-468-576B-17	Sequence 17, Appl

28	36	42.4	1745	2 US-09-031-485-33	Sequence 33, Appl
29	36	42.4	1745	2 US-08-847-429A-33	Sequence 33, Appl
30	36	42.4	1745	3 US-09-065-474-33	Sequence 33, Appl
31	35.5	41.8	113	4 US-09-180-077-6	Sequence 6, Appl
32	35.5	41.8	113	4 US-09-180-077-11	Sequence 11, Appl
33	35	41.2	311	2 US-08-775-009-32	Sequence 32, Appl
34	35	41.2	311	2 US-08-775-009-33	Sequence 33, Appl
35	35	41.2	353	4 US-09-118-442-6	Sequence 6, Appl
36	35	41.2	353	4 US-09-677-064-6	Sequence 6, Appl
37	35	41.2	2296	2 US-08-286-819A-27	Sequence 27, Appl
38	35	41.2	2296	3 US-08-980-357-27	Sequence 27, Appl
39	34	40.0	339	1 US-08-626-994A-3	Sequence 3, Appl
40	34	40.0	339	3 US-08-957-742-3	Sequence 3, Appl
41	34	40.0	364	1 US-08-626-994A-1	Sequence 1, Appl
42	34	40.0	364	3 US-08-957-742-1	Sequence 1, Appl
43	34	40.0	375	2 US-08-446-875-8	Sequence 8, Appl
44	34	40.0	375	2 US-08-102-385G-8	Sequence 8, Appl
45	34	40.0	432	2 US-08-700-152A-4	Sequence 4, Appl
46	34	40.0	519	2 US-08-725-736D-2	Sequence 2, Appl
47	34	40.0	519	3 US-09-162-368B-2	Sequence 2, Appl
48	34	40.0	519	4 US-09-161-877B-2	Sequence 2, Appl
49	34	40.0	671	6 5266464-2	Patent No. 5266464
50	34	40.0	685	3 US-09-031-563-21	Sequence 21, Appl

ALIGNMENTS

RESULT 1
US-07-603-133B-13
Sequence 13, Application US/07603133B
Patent No. 5298244
GENERAL INFORMATION:
APPLICANT: Redmond, Mark J.
APPLICANT: Ijaz, Mohammed K.
TITLE OF INVENTION: ASSEMBLED VIRAL PARTICLES AND THEIR
NUMBER OF SEQUENCES: 30
TITLE OF INVENTION: USE IN A VACCINE TO ROTAVIRAL DISEASE
CORRESPONDENCE ADDRESS:
ADDRESSEE: Morrison & Foerster
STREET: 545 Middlefield Road, Suite 200
CITY: Menlo Park
STATE: CA
COUNTRY: USA
ZIP: 94025
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/603.133B
FILING DATE: 19901025
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Robins, Roberta L.
REGISTRATION NUMBER: 33,208
REFERENCE/DOCKET NUMBER: 9313-0004.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 327-7250
TELEFAX: (415) 327-2951
TELEX: 706141
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 775 amino acids
TYPE: AMINO ACID
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-07-603-133B-13

Query Match 49.4%; Score 42; DB 1; Length 775;
Best Local Similarity 43.8%; Pred. No. 28;
Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;
OY 1 NMLVPMNPLVNMNEFN 16
DB 741 NMLKSNPNVLRNFTN 756

RESULT 2
US-07-603-133B-14
Sequence 14 Application US/07603133B
Patent No. 5298244
GENERAL INFORMATION:
APPLICANT: Redmond, Mark J.
APPLICANT: Ijaz, Mohammed K.
APPLICANT: Parker, Michael D.
TITLE OF INVENTION: ASSEMBLED VIRAL PARTICLES AND THEIR
TITLE OF INVENTION: USE IN A VACCINE TO ROTAVIRAL DISEASE
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESSEE: Morrison & Foerster
STREET: 545 Middlefield Road, Suite 200
CITY: Menlo Park
STATE: CA
COUNTRY: USA
ZIP: 94025
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/603,133B
FILING DATE: 19901025
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Robins, Roberta L.
REGISTRATION NUMBER: 33,208
REFERENCE/DOCKET NUMBER: 9313-0004.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 327-7250
TELEFAX: (415) 327-2951
TELEX: 706141
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 775 amino acids
TYPE: AMINO ACID
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-07-603-133B-14
Query Match 47.1%; Score 40; DB 1; Length 775;
Best Local Similarity 46.7%; Pred. No. 60;
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
OY 2 NMLVPMNPLVNMNEFN 16
DB 742 NMLKSNPNVLRNFTN 756

NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESSEE: Morrison & Foerster
STREET: 545 Middlefield Road, Suite 200
CITY: Menlo Park
STATE: CA
COUNTRY: USA
ZIP: 94025
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/603,133B
FILING DATE: 19901025
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Robins, Roberta L.
REGISTRATION NUMBER: 33,208
REFERENCE/DOCKET NUMBER: 9313-0004.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 327-7250
TELEFAX: (415) 327-2951
TELEX: 706141
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 776 amino acids
TYPE: AMINO ACID
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-07-603-133B-17

Query Match 47.1%; Score 40; DB 1; Length 776;
Best Local Similarity 43.8%; Pred. No. 60;
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;
OY 1 NMLVPMNPLVNMNEFN 16
DB 742 NMLKSNPNVLRNFTN 757

RESULT 4
US-07-603-133B-20
Sequence 20 Application US/07603133B
Patent No. 5298244
GENERAL INFORMATION:
APPLICANT: Redmond, Mark J.
APPLICANT: Ijaz, Mohammed K.
APPLICANT: Parker, Michael D.
TITLE OF INVENTION: ASSEMBLED VIRAL PARTICLES AND THEIR
TITLE OF INVENTION: USE IN A VACCINE TO ROTAVIRAL DISEASE
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESSEE: Morrison & Foerster
STREET: 545 Middlefield Road, Suite 200
CITY: Menlo Park
STATE: CA
COUNTRY: USA
ZIP: 94025
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/603,133B
FILING DATE: 19901025
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Robins, Roberta L.

PI Sabara MIT, Frenchick PJ, Potter AA, Ijaz MK, Gilchrist JE;
XX WPI: 1987-228567/33.
XX
PT New peptide fragments of rotaviral proteins - useful conjugates
PT In vaccines for protecting against gastrointestinal disorders and
PT diarrhoea
PS Disclosure: Fig 3; 84pp; English.
XX
CC The peptide fragments of glycoprotein VP7 and proteins VP6 and VP3
CC of rotaviruses are useful when attached to carriers as vaccines for
CC birds and mammals, including man. The vaccines confer protection
CC against gastrointestinal disorders and diarrhoea produced by the
CC rotaviruses. For use in vaccines the peptides are covalently linked
CC to eg, keyhole limpet haemocyanin, BSA, ovalbumin, poly-L-lysine,
CC or VP6 bovine rotavirus protein. An adjuvant may be included.

Sequence 776 AA;

Query Match 47.1%; Score 40; DB 8; Length 776;
Best Local Similarity 43.8%; Pred. No. 99;
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

OY 1 MNLVPMNPVNMNPN 16
:11:11:11:11:
DB 742 Inlrrsdprvrlrefln 757

RESULT 2

AAR24293
ID AAR24293 standard; Protein; 776 AA.

AC AAR24293;

DT 18-NOV-1992 (first entry)

DE VP4 protein of strain C486 (bovine).

KW Vaccine: rotavirus; viral protein; virus; capsid;

KW VP6; VP4; VP7.

OS Bovine C486 rotavirus.

PN WC9207941-A.

XX 14-MAY-1992.

PF 24-OCT-1991; 91MO-CA00376.

PR 25-OCT-1990; 90US-0603133.

PA (UYSA-) UNIV SASKATCHEWAN.

PI Ijaz MK, Parker MD, Redmond MJ;

XX WPI: 1992-183676/22.

DR N-PSDB; AAQ25167.

PT Assembled viral particles useful as a rota-viral vaccine -
PT contain the inner capsid protein VP6 in combination with either
PT or both of the outer capsid proteins VP4 and VP7

PS Disclosure: Fig 4(1-9); 103pp; English.

CC Assembled viral particles include the inner capsid protein, VP6,
CC in combination with either or both of the outer capsid proteins,
CC VP4 and VP7. These assemblies can be used in vaccine compns.
CC for the treatment and prevention of rotaviral disease.

Sequence 776 AA;

Query Match 47.1%; Score 40; DB 13; Length 776;
Best Local Similarity 43.8%; Pred. No. 99;
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

OY 1 MNLVPMNPVNMNPN 16
:11:11:11:11:
DB 742 Inlrrsdprvrlrefln 757

RESULT 3

AAB07418
ID AAB07418 standard; protein; 776 AA.

AC AAB07418;

DT 20-OCT-2000 (first entry)

DE Amino acid sequence of a viral protein 7 (VP7) of rotavirus.

KW Immunochemical; rotavirus; VP6; viral protein; VP4; antibody response;

KW vaccine; rotavirus infection.

OS Rotavirus.

PN US6086880-A.

PD 11-JUL-2000.

PF 07-JUL-1993; 93US-0089397.

PR 03-SEP-1986; 86US-0903325.

PR 07-SEP-1988; 88US-0241761.

PR 27-FEB-1991; 91US-0661859.

PR 26-DEC-1985; 85US-0813661.

PR 12-JUL-1990; 90US-0552350.

PR 10-DEC-1990; 90US-0626041.

PA (UYSA-) UNIV SASKATCHEWAN.

PI Sabara MIT, Frenchick PJ, Potter AA, Gilchrist JE, Redmond MJ;

XX Ijaz MK;

XX WPI: 2000-498191/44.

PT Peptide useful as vaccine for protection against infection by rotavirus

PT comprises specified subunits of rotavirus VP6 and VP4 viral proteins -

PS Disclosure: Fig 1; 34pp; English.

CC The specification describes peptides which are useful as

CC immunochemicals. The peptides comprise subunits 40-60 of rotavirus VP6

CC viral protein, 232-255 of VP4 (both optionally modified to facilitate

CC binding to carrier), where one or more of valine residues of VP4 is

CC substituted at positions 234, 236 or 235 by alanine, or subunits 240-248

CC of VP4 which is optionally modified to facilitate covalent coupling to a

CC carrier other than VP6. Rotavirus VP4 subunit was found to be capable of

CC plaque reduction when mixed with infection virus. The peptides are

CC useful for initiating an antibody response in a mammal against

CC rotaviruses. The peptides are also useful as vaccines to protect against

CC rotavirus infection and as diagnostic tools to detect the presence of

CC rotaviral infection. VP4 subunits are useful for prophylactic protection

CC or for therapy with respect to rotaviral infection. AAB07553-55 and

CC AAB07418 represent rotavirus VP7 proteins.

Sequence 776 AA;

Query Match 47.1%; Score 40; DB 21; Length 776;
Best Local Similarity 43.8%; Pred. No. 99;
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

OY 1 MNLVPMNPVNMNPN 16

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 6, 2002, 09:33:20 ; Search time 135.97 Seconds
(without alignments)
17.212 Million cell updates/sec

Title: US-09-897-042-18
Perfect score: 85
Sequence: 1 MNLVPMNPLVMEFPM 16

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

imm DB seq length: 0
imm DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database :
1: SP_ARCHAEA:***
2: SP_BACTERIA:***
3: SP_FUNGI:***
4: SP_HUMAN:***
5: SP_INVERTEBRATE:***
6: SP_MAMMAL:***
7: SP_MHC:***
8: SP_ORGANELLE:***
9: SP_PHAGE:***
10: SP_PLANT:***
11: SP_PROTOIST:***
12: SP_VIRUS:***
13: SP_VERTEBRATE:***
14: SP_UNCLASSIFIED:***

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Hit No.	Score	Query Match	Length	DB ID	Description
1	44	51.8	775	12 O9WAL4	O9WAL4 human rotav
2	44	51.8	775	12 O9WAL1	O9WAL1 human rotav
3	44	51.8	775	12 O9WAL0	O9WAL0 human rotav
4	44	51.8	775	12 O9WAK8	O9WAK8 human rotav
5	44	51.8	775	12 O9WAK4	O9WAK4 human rotav
6	44	51.8	775	12 O9WAK3	O9WAK3 human rotav
7	44	51.8	775	12 O9WAK2	O9WAK2 human rotav
8	44	51.8	775	12 O9WAK0	O9WAK0 human rotav
9	44	51.8	775	12 O9WAK0	O9WAK0 human rotav
10	43	50.6	179	8 O9GAG5	O9GAG5 polyrhachis
11	43	50.6	500	10 O9S9W4	O9S9W4 arabidopsis
12	42	48.4	226	8 O9G2G2	O9G2G2 phytomyza n
13	42	48.4	319	8 O9B6G2	O9B6G2 thryonoms
14	42	48.4	330	6 O9BFF1	O9BFF1 sytilagus
15	42	48.4	330	6 O9BFF0	O9BFF0 ochotona hy
16	42	48.4	346	2 O9RVT6	O9RVT6 deinococcus
17	42	48.4	775	12 O82012	O82012 human rotav
18	42	48.4	775	12 O82118	O82118 human rotav
19	42	48.4	775	12 O86111	O86111 rotavirus g

20	42	49.4	775	12 O86185	O86185 rotavirus S
21	42	49.4	775	12 O9WAL7	O9WAL7 human rotav
22	42	49.4	775	12 O9WAL6	O9WAL6 human rotav
23	42	49.4	775	12 O9WAL3	O9WAL3 human rotav
24	42	49.4	775	12 O9WAK9	O9WAK9 human rotav
25	42	49.4	775	12 O9WAK5	O9WAK5 human rotav
26	42	49.4	775	12 O9WAK1	O9WAK1 human rotav
27	42	49.4	775	12 O9E3V4	O9E3V4 human rotav
28	42	49.4	775	12 O9DUU6	O9DUU6 human rotav
29	42	49.4	775	12 O99IN3	O99IN3 human rotav
30	42	49.4	775	12 O99IN2	O99IN2 human rotav
31	42	49.4	775	12 O99IN1	O99IN1 human rotav
32	42	49.4	775	12 O99IN0	O99IN0 human rotav
33	42	49.4	775	12 O99IM9	O99IM9 human rotav
34	42	49.4	775	12 O99IM8	O99IM8 human rotav
35	42	49.4	775	12 O99IM7	O99IM7 human rotav
36	42	49.4	775	12 O98VL3	O98VL3 human rotav
37	41	48.2	179	8 O9GAK8	O9GAK8 oecophylla
38	41	48.2	184	8 O03164	O03164 greya solen
39	41	48.2	185	8 O36040	O36040 tridentator
40	41	48.2	576	2 O9CIN6	O9CIN6 pasteurilla
41	41	48.2	620	5 O16519	O16519 caenorhabd
42	41	48.2	994	5 O77070	O77070 placopecten
43	40	47.1	100	8 O9G2E1	O9G2E1 hyloms sui
44	40	47.1	145	8 O9G7P3	O9G7P3 diaseopsis
45	40	47.1	153	10 O9M2J3	O9M2J3 arabidopsis
46	40	47.1	159	8 O34592	O34592 greya mitel
47	40	47.1	170	8 O34614	O34614 greya obscu
48	40	47.1	174	8 O9TB29	O9TB29 greya solen
49	40	47.1	179	8 O9GAG8	O9GAG8 polyrhachis
50	40	47.1	180	8 O34562	O34562 greya enchr

ALIGNMENTS

RESULT 1
ID O9WAL4 PRELIMINARY: PRT: 775 AA.
AC O9WAL4;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE VP4.
OS Human rotavirus 3.
OC Viruses; dsRNA viruses; Reoviridae; Rotavirus.
OX NCBI_TaxID=36430;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TK-28;
RX MEDLINE=99244278; PubMed=10229272;
RA Cao X.R., Akibara S., Pang Z.Y., Nakagomi O., Ushijima H.;
RT "Genetic variation in the VP4 and NSP4 genes of human rotavirus
serotype 3 (G3 type) isolated in China and Japan.";
RL Microbiol. Immunol. 43:171-175(1999).
DR EMBL: AB008288; BAA77553.1;
DR InterPro: IPR000416; Cap_VP4.
DR Pfam: PF00426; VP4; 1.
SO SEQUENCE 775 AA; 87276 MW; B4DBD22DFD7CCED6 CRC64;

Query Match 51.8%; Score 44; DB 12; Length 775;
Best Local Similarity 43.8%; Pred. No. 32;
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

OY 1 MNLVPMNPLVMEFPM 16
Db 741 LNDKSNPVLRFDFIN 756

RESULT 2
ID O9WAL1 PRELIMINARY: PRT: 775 AA.

AC Q9WAL1;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE VP4.
OS Human rotavirus 3.
OC Viruses; dsRNA viruses; Reoviridae; Rotavirus.
OX NCBI_TaxID=36430;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-AI-75;
RX MEDLINE=99244278; PubMed=10229272;
RA Cao X.R., Akihara S., Fang Z.Y., Nakagomi O., Ushijima H.;
RT "Genetic variation in the VP4 and NSP4 genes of human rotavirus
serotype 3 (G3 type) isolated in China and Japan.";
RL Microbiol. Immunol. 43:171-175(1999).
DR EMBL: AB008285; BAA77550.1; -;
DR InterPro: IPR000416; Cap_VP4.
DR Pfam: PF00426; VP4; 1.
SQ SEQUENCE 775 AA; 87408 MW; AAI451628B4F9EB CRC64;

Query Match 51.8%; Score 44; DB 12; Length 775;
Best Local Similarity 43.8%; Pred. No. 32;
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 1 MNLVPMNPVLMNEFMN 16
Db 741 LNLIKSNPNVLRDFIN 756

RESULT 3
Q9WAL0
ID Q9WAL0; PRELIMINARY; PRT; 775 AA.
AC Q9WAL0;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE VP4.
OS Human rotavirus 3.
OC Viruses; dsRNA viruses; Reoviridae; Rotavirus.
OX NCBI_TaxID=36430;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-AI-53;
RX MEDLINE=99244278; PubMed=10229272;
RA Cao X.R., Akihara S., Fang Z.Y., Nakagomi O., Ushijima H.;
RT "Genetic variation in the VP4 and NSP4 genes of human rotavirus
serotype 3 (G3 type) isolated in China and Japan.";
RL Microbiol. Immunol. 43:171-175(1999).
DR EMBL: AB008284; BAA77549.1; -;
DR InterPro: IPR000416; Cap_VP4.
DR Pfam: PF00426; VP4; 1.
SQ SEQUENCE 775 AA; 87322 MW; F9C2C714320A183D CRC64;

Query Match 51.8%; Score 44; DB 12; Length 775;
Best Local Similarity 43.8%; Pred. No. 32;
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 1 MNLVPMNPVLMNEFMN 16
Db 741 LNLIKSNPNVLRDFIN 756

RESULT 4
Q9WAK8
ID Q9WAK8; PRELIMINARY; PRT; 775 AA.
AC Q9WAK8;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE VP4.

OS Human rotavirus 3.
OC Viruses; dsRNA viruses; Reoviridae; Rotavirus.
OX NCBI_TaxID=36430;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-AI-36;
RX MEDLINE=99244278; PubMed=10229272;
RA Cao X.R., Akihara S., Fang Z.Y., Nakagomi O., Ushijima H.;
RT "Genetic variation in the VP4 and NSP4 genes of human rotavirus
serotype 3 (G3 type) isolated in China and Japan.";
RL Microbiol. Immunol. 43:171-175(1999).
DR EMBL: AB008282; BAA77547.1; -;
DR InterPro: IPR000416; Cap_VP4.
DR Pfam: PF00426; VP4; 1.
SQ SEQUENCE 775 AA; 87309 MW; 9FB8793A8BC5B38 CRC64;

Query Match 51.8%; Score 44; DB 12; Length 775;
Best Local Similarity 43.8%; Pred. No. 32;
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 1 MNLVPMNPVLMNEFMN 16
Db 741 LNLIKSNPNVLRDFIN 756

RESULT 5
Q9WAK4
ID Q9WAK4; PRELIMINARY; PRT; 775 AA.
AC Q9WAK4;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE VP4.
OS Human rotavirus 3.
OC Viruses; dsRNA viruses; Reoviridae; Rotavirus.
OX NCBI_TaxID=36430;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MO;
RX MEDLINE=99244278; PubMed=10229272;
RA Cao X.R., Akihara S., Fang Z.Y., Nakagomi O., Ushijima H.;
RT "Genetic variation in the VP4 and NSP4 genes of human rotavirus
serotype 3 (G3 type) isolated in China and Japan.";
RL Microbiol. Immunol. 43:171-175(1999).
DR EMBL: AB008278; BAA77543.1; -;
DR InterPro: IPR000416; Cap_VP4.
DR Pfam: PF00426; VP4; 1.
SQ SEQUENCE 775 AA; 87288 MW; E4CE723A8124818F CRC64;

Query Match 51.8%; Score 44; DB 12; Length 775;
Best Local Similarity 43.8%; Pred. No. 32;
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 1 MNLVPMNPVLMNEFMN 16
Db 741 LNLIKSNPNVLRDFIN 756

RESULT 6
Q9WAK3
ID Q9WAK3; PRELIMINARY; PRT; 775 AA.
AC Q9WAK3;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE VP4.
OS Human rotavirus 3.
OC Viruses; dsRNA viruses; Reoviridae; Rotavirus.
OX NCBI_TaxID=36430;
RN [1]
RP SEQUENCE FROM N.A.

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 6, 2002, 09:34:42 ; Search time 68.88 Seconds
(without alignments)
8.517 Million cell updates/sec

Title: US-09-897-042-18

Perfect score: 85

Sequence: 1 MNLVPMNPLVNMERN 16

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues
Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	43	50.6	2054	1 YCF2_PINTH	P16533 pinus thunb
2	42	49.4	775	1 VP4_ROTTH	P13842 human rotav
3	42	49.4	775	1 VP4_ROTTH	P21284 human rotav
4	41	48.2	241	1 YF45_METJA	O58940 methanococ
5	41	48.2	576	1 CYDC_HAEIN	P45081 haemophilus
6	40	47.1	184	1 21KD_ZYMO	P15256 zymomonas m
7	40	47.1	334	1 Y009_BORBU	O51042 borrelia bu
8	40	47.1	775	1 VP4_NCDV	P17465 nebraska ca
9	40	47.1	775	1 VP4_ROTTH	P30214 human rotav
10	40	47.1	775	1 VP4_ROTTH	P11196 human rotav
11	40	47.1	775	1 VP4_ROTTH	P11195 human rotav
12	40	47.1	775	1 VP4_ROTTH	P11193 human rotav
13	40	47.1	776	1 VP4_ROTTH	P04508 simian 11 r
14	40	47.1	776	1 VP4_ROTTH	P36305 bovine rota
15	40	47.1	776	1 VP4_ROTTH	P08713 bovine rota
16	40	47.1	776	1 VP4_ROTTH	P17463 simian 11 r
17	40	47.1	908	1 H104_YEAST	P31539 saccharomyc
18	39	45.9	536	1 CC44_DROME	O99600 drosophila
19	39	45.9	590	1 NUSM_DINSE	O79556 dinodon sem
20	39	45.9	775	1 VP4_ROTTH	P11194 human rotav
21	39	45.9	776	1 VP4_ROTTH	O08778 human rotav
22	39	45.9	1018	1 VGNM_BPMV	P23009 bean-pod mo
23	38	44.7	200	1 ATKC_ANASL	O97649 anabaena sp
24	38	44.7	314	1 MHPA_ECOLI	P54711 escherichia
25	38	44.7	348	1 OPSP_GLOME	O62782 globicephal
26	38	44.7	413	1 EF1A_HELVI	P55276 heliothis v
27	38	44.7	463	1 EF1A_BOMO	P29520 bombyx mori
28	38	44.7	775	1 VP4_ROTTH	P39033 feline rota
29	38	44.7	775	1 VP4_ROTTH	P39033 human rotav
30	38	44.7	775	1 VP4_ROTTH	O01641 human rotav
31	38	44.7	776	1 VP4_ROTTH	P12976 simian 11 r
32	38	44.7	776	1 VP4_ROTTH	O06894 canine rota
33	38	44.7	776	1 VP4_ROTTH	O02945 equine rota

ALIGNMENTS

34	38	44.7	776	1 VP4_ROTTH	O07416 feline rota
35	38	44.7	776	1 VP4_ROTTH	O06895 human rotav
36	38	44.7	776	1 VP4_ROTTH	P12473 rhesus rota
37	38	44.7	776	1 VP4_ROTTH	P17464 simian 11 r
38	38	44.7	1142	1 SPK_HUMAN	O92797 homo sapien
39	38	44.7	2179	1 POLG_EC23W	O73556 e genome po
40	37	43.5	284	1 STCI_YEAST	P38634 saccharomyc
41	37	43.5	347	1 NUD2M_PIG	O79875 sus scrofa
42	37	43.5	379	1 CYB_LOCOMI	O36427 locusta mig
43	37	43.5	389	1 BMRI_BACSU	P33449 bacillus su
44	37	43.5	439	1 EF12_EUPCR	O27140 euphorbia cr
45	37	43.5	447	1 EF12_DAUCA	P34823 daucus caro
46	37	43.5	463	1 EF11_DROME	P08736 drosophila
47	37	43.5	536	1 YC42_SYNY3	P42349 synchocyst
48	37	43.5	712	1 RY3_CHLRE	O08365 chlamydomon
49	37	43.5	729	1 KAR3_YEAST	P17119 saccharomyc
50	37	43.5	765	1 TRRC_PHYBL	P20409 p anthranil

RESULT 1
YCF2_PINTH STANDARD: PRT; 2054 AA.
ID YCF2_PINTH
AC P16533;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL 244.6 KDA PROTEIN YCF2 (ORF 2054).
GN YCF2.
OS Pinus thunbergii (Green pine) (Japanese black pine).
OG Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus.
OX NCBI_TaxID=3350;
RN [1]
RP MEDLINE=95024047; PubMed=7937893;
RA Wakasugi T., Tsudzuki J., Ito S., Nakashima K., Tsudzuki T.,
Sugliura M.,
RT "Loss of all ndh genes as determined by sequencing the entire
RT chloroplast genome of the black pine Pinus thunbergii".
RT Proc. Natl. Acad. Sci. U.S.A. 91:9794-9798(1994).
CC -!- FUNCTION: NOT YET KNOWN.
CC -!- SIMILARITY: BELONGS TO THE YCF2 FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: D17510; BAA04460.1;
DR InterPro: IPR001939; AAA_subfam.
DR Pfam: PF00004; AAA_1.
KW Chloroplast; Hypothetical protein.
SQ SEQUENCE 2054 AA; 244604 MW; 6F5E9ZD078E33A9A CRC64;

Query Match 50.6%; Score 43; DB 1; Length 2054;
Best Local Similarity 46.7%; Pred. No. 34;
Matches 7; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 2 MNLVPMNPLVNMERN 16
I:|||||:|
Db 717 NKIRINPLIDFDN 731

RESULT 2
VP4_ROTTH

ID VP4_ROTHER STANDARD: PRT: 775 AA.
AC P13842;
DT 01-JAN-1990 (Rel. 13, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE OUTER CAPSID PROTEIN VP4 (HEMAGGLUTININ) (OUTER LAYER PROTEIN VP4)
GN [CONTAINS: OUTER CAPSID PROTEINS VP5 AND VP8].
OS Human rotavirus (strain KU).
OC Viruses; dsRNA viruses; Reoviridae; Rotavirus.
OX NCBI_TaxID=10952;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88230603; PubMed=2453680;
RA Taniguchi K., Maloy W.L., Nishikawa K., Green K.Y., Hoshino Y.,
Urasawa S., Kapikian A.Z., Chanock R.M., Gorziglia M.;
RT "Identification of cross-reactive and serotype 2-specific
neutralization epitopes on VP3 of human rotavirus.";
J. Virol. 62:2421-2426(1988).
CC -1- SUBCELLULAR LOCATION: OUTER CAPSID.
CC -1- PTM: VP8 IS ONE OF TWO TRYPSIN CLEAVAGE PRODUCTS OF VP4; THE
OTHER PRODUCT IS VP5.
CC -1- SIMILARITY: BELONGS TO THE ROTAVIRUS VP4 PROTEIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M21014; AAA47334.1; -
DR PIR: A28844; VPRHR.
DR InterPro: IPR000416; Cap_VP4.
DR Pfam: PF00426; VP4; 1.
KW Coat protein; Glycoprotein.
FT CHAIN 1 775 OUTER CAPSID PROTEIN VP4.
FT CHAIN 1 240 OUTER CAPSID PROTEIN VP8.
FT CHAIN 247 775 OUTER CAPSID PROTEIN VP5.
FT CARBOHYD 32 32 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 56 56 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 97 97 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 132 132 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 324 324 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 583 583 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 589 589 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 592 592 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 599 599 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 775 AA; 87538 MW; 3222AE197D0DFD CRC64;
Query Match 49.4%; Score 42; DB 1; Length 775;
Best Local Similarity 43.8%; Pred. No. 17;
Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;
QY 1 MNLVPMNPVMEFMN 16
Db 741 LNLKSNPNVLRNFN 756
RESULT 3
ID VP4_ROTHER STANDARD: PRT: 775 AA.
AC P21284;
DT 01-MAY-1991 (Rel. 18, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE OUTER CAPSID PROTEIN VP4 (HEMAGGLUTININ) (OUTER LAYER PROTEIN VP4)
GN [CONTAINS: OUTER CAPSID PROTEINS VP5 AND VP8].
OS Human rotavirus (strain L26).
OC Viruses; dsRNA viruses; Reoviridae; Rotavirus.

OX NCBI_TaxID=10953;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91012813; PubMed=2170690;
RA Taniguchi K., Urasawa T., Kobayashi N., Gorziglia M., Urasawa S.;
RT "Nucleotide sequence of VP4 and VP7 genes of human rotaviruses with
subgroup I specificity and long RNA pattern: implication for new G
serotype specificity.";
J. Virol. 64:5640-5644(1990).
CC -1- SUBCELLULAR LOCATION: OUTER CAPSID.
CC -1- PTM: VP8 IS ONE OF TWO TRYPSIN CLEAVAGE PRODUCTS OF VP4; THE
OTHER PRODUCT IS VP5.
CC -1- SIMILARITY: BELONGS TO THE ROTAVIRUS VP4 PROTEIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M58292; AAA47335.1; -
DR PIR: A36410; VPRRWL.
DR PIR: C36410; VPRRWL.
DR InterPro: IPR000416; Cap_VP4.
DR Pfam: PF00426; VP4; 1.
KW Coat protein; Glycoprotein.
FT CHAIN 1 775 OUTER CAPSID PROTEIN VP4.
FT CHAIN 1 240 OUTER CAPSID PROTEIN VP8.
FT CHAIN 247 775 OUTER CAPSID PROTEIN VP5.
FT CARBOHYD 32 32 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 56 56 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 97 97 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 132 132 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 150 150 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 195 195 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 324 324 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 583 583 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 589 589 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 599 599 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARIANT 51 51 R -> G (IN STRAIN L27).
FT VARIANT 392 392 E -> C (IN STRAIN L27).
FT VARIANT 405 405 F -> C (IN STRAIN L27).
SQ SEQUENCE 775 AA; 87650 MW; 40CA498305DEC63 CRC64;
Query Match 49.4%; Score 42; DB 1; Length 775;
Best Local Similarity 43.8%; Pred. No. 17;
Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;
QY 1 MNLVPMNPVMEFMN 16
Db 741 LNLKSNPNVLRNFN 756
RESULT 4
ID YF45_METJA STANDARD: PRT: 241 AA.
AC 058940;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL PROTEIN M1545.
GN M1545.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 6, 2002, 09:16:49 ; Search time 77.97 Seconds

(without alignments)
15.632 Million cell updates/sec

Title: US-09-897-042-18

Perfect score: 85

Sequence: 1 MNLVPMNPLVNMNFMN 16

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 50 summaries

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	43	50.6	500	2 G85069	hypothetical prote
2	43	50.6	2054	2 T07584	hypothetical prote
3	42	49.4	346	2 F75457	conserved hypotet
4	42	49.4	775	1 VPXRHK	outer layer protei
5	42	49.4	775	1 VPXRML	outer layer protei
6	42	49.4	775	1 VPXRMM	outer layer protei
7	42	49.4	775	2 S52165	outer capsid prote
8	42	49.4	776	2 A48480	outer capsid prote
9	41	48.2	241	2 H64492	hypothetical prote
10	41	48.2	576	2 E64186	probable ATP-bindi
11	41	48.2	620	2 T31906	hypothetical prote
12	40	47.1	153	2 T45983	hypothetical prote
13	40	47.1	184	2 S06695	hypothetical prote
14	40	47.1	204	2 T03962	hypothetical prote
15	40	47.1	334	2 A70101	hypothetical prote
16	40	47.1	362	2 T33502	hypothetical prote
17	40	47.1	382	2 E85082	hypothetical prote
18	40	47.1	382	2 T14186	hypothetical prote
19	40	47.1	656	2 T52064	deaf-like protein
20	40	47.1	775	1 VPXRK2	outer layer protei
21	40	47.1	775	1 VPXRK3	outer layer protei
22	40	47.1	775	1 VPXRK4	outer layer protei
23	40	47.1	775	1 VPXRK5	outer layer protei
24	40	47.1	775	1 VPXRK6	outer layer protei
25	40	47.1	776	1 VPXRK7	outer layer protei
26	40	47.1	776	1 VPXRK8	outer layer protei
27	40	47.1	776	1 T02022	outer layer protei
28	40	47.1	776	2 S24410	hypothetical prote
29	40	47.1	908	1 S61476	endopeptidase Clp

30	39.5	46.5	486	2 S31805	VP5 protein - porc
31	39.5	46.5	2910	2 T28156	DNA-directed RNA p
32	39	45.9	333	2 E86444	hypothetical prote
33	39	45.9	334	2 T46238	hypothetical prote
34	39	45.9	343	2 T33945	hypothetical prote
35	39	45.9	427	2 A84820	hypothetical prote
36	39	45.9	461	2 T51991	translational elonga
37	39	45.9	590	2 T11098	MDH dehydrogenase
38	39	45.9	775	1 VPXRK6	outer layer protei
39	39	45.9	799	2 F83549	probable ATP-depen
40	39	45.9	1018	1 GNMKG7	genome polypeptid
41	39	45.9	1997	2 F71607	DNA helicase II BR
42	39	45.9	2254	2 D86215	protein T6D22.14 l
43	38.5	45.3	419	2 T23948	hypothetical prote
44	38.5	45.3	883	2 T23948	hypothetical prote
45	38	44.7	63	2 T12121	MDH dehydrogenase
46	38	44.7	141	2 S08523	hypothetical prote
47	38	44.7	200	2 T46848	K ⁺ -transporting AT
48	38	44.7	314	2 D64762	2,3-dihydroxypheny
49	38	44.7	314	2 G85529	2,3-dihydroxypheny
50	38	44.7	343	2 G72218	conserved hypotet

ALIGNMENTS

RESULT 1

G85069

hypothetical protein AT4g05550 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 16-Feb-2001

C:Accession: G85069

R:Anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Sp

Nature 402, 769-777, 1999

A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.

A:Reference number: A85001; MUID:20083488

A:Accession: G85069

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-500 <STO>

A:Cross-references: GB:NC_001268; MID:g7267315; PIDN:CAB81097.1; GSPDB:GN00140

C:Genetics:

A:Gene: AT4g05550

A:Map position: 4

Query Match 50.6%; Score 43; DB 2; Length 500;

Best Local Similarity 50.0%; Pred. NO. 15;

Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY 1 MNLVPMNPLVNMNFMN 16

Db 1 MNLSTLEDIVLKEFMN 16

RESULT 2

T07584

hypothetical protein 2054 - Japanese black pine chloroplast

C:Species: chloroplast Pinus thunbergiana (Japanese black pine)

C:Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 18-Aug-2000

C:Accession: T07584

R:Wakasugi, T.; Tsudzuki, J.; Ito, S.; Nakashima, K.; Tsudzuki, T.; Sugiyama, M.

Proc. Natl. Acad. Sci. U.S.A. 91, 9794-9798, 1994

A:Title: Loss of all ndh genes as determined by sequencing the entire chloroplast gen

A:Reference number: Z16030; MUID:95024047

A:Accession: T07584

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-2054 <WAK>

A:Cross-references: EMBL:DJ7510; NID:g529643; PIDN:BA04460.1; PID:g1262745

C:Genetics:

A:Gene: chloroplast

C:Keywords: chloroplast

Query Match 50.6%; Score 43; DB 2; Length 2054;
Best Local Similarity 46.7%; Pred. No. 77;
Matches 7; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

OY 2 NMLVPMNPVLMNEFMN 16
DB 717 NKLIPNPLIDFEDN 731

RESULT 3

F75457
Conserved hypothetical protein - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C:Accession: F75457
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
M.; Shen, M.; Vamathavan, J.J.; Lam, P.; McDonald, L.; Ulterback, T.; Zalewski, C.; Ma
.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896
A:Accession: F75457
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-346 <MID>
A:Cross-references: GB:AE001946; GB:AE000513; NID:g6458655; PIDN:AAF10513.1; PID:g645866
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR0936
A:Map position: 1
C:Superfamily: conserved hypothetical protein H10365

Query Match 49.4%; Score 42; DB 2; Length 346;
Best Local Similarity 43.8%; Pred. No. 14;
Matches 7; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

OY 1 NMLVPMNPVLMNEFMN 16
DB 279 VNLIPMNPWDGSDFFVS 294

RESULT 4

VPXRK
outer layer protein VP3 - human rotavirus A (strain KU)
N:Alternate names: glycoprotein VP3; hemagglutinin; outer capsid protein VP3
C:Species: human rotavirus A
C:Date: 30-Sep-1989 #sequence_revision 30-Sep-1989 #text_change 16-Jul-1999
C:Accession: A28844
R:Taniguchi, K.; Maloy, W.L.; Nishikawa, K.; Green, K.Y.; Hoshino, Y.; Urasawa, S.; Kapl
J. Virol. 62, 2421-2426, 1988
A:Title: Identification of cross-reactive and serotype 2-specific neutralization epitope
A:Reference number: A28844; MUID:88230603
A:Accession: A28844
A:Molecule type: mRNA
A:Residues: 1-775 <TRAN>
A:Cross-references: GB:M21014; NID:g333852; PIDN:AAA47334.1; PID:g333853
C:Genetics:
A:Map position: segment 4
C:Superfamily: rotavirus outer layer protein VP3
C:Keywords: glycoprotein; hemagglutinin; outer capsid protein
F:32,56,97,132,324,583,589,592,599/Binding site: carbohydrate (Asn) (covalent) #status F

Query Match 49.4%; Score 42; DB 1; Length 775;
Best Local Similarity 43.8%; Pred. No. 36;
Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

OY 1 NMLVPMNPVLMNEFMN 16
DB 741 LNLIKSNPNVLRNFN 756

RESULT 5
VPXRK
outer layer protein VP3 - human rotavirus A (strain L26)
N:Alternate names: VP4 protein
C:Species: human rotavirus A
C:Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 16-Jul-1999
C:Accession: A36410
R:Taniguchi, K.; Urasawa, T.; Kobayashi, N.; Gorziglia, M.; Urasawa, S.
J. Virol. 64, 5640-5644, 1990
A:Title: Nucleotide sequence of VP4 and VP7 genes of human rotaviruses with subgroup
A:Reference number: A36410; MUID:91012813
A:Accession: A36410
A:Molecule type: genomic RNA
A:Residues: 1-775 <TRAN>
A:Cross-references: EMBL:M58292; NID:g333854; PIDN:AAA47335.1; PID:g333855
C:Genetics:
A:Map position: segment 4
C:Superfamily: rotavirus outer layer protein VP3
C:Keywords: capsid protein; glycoprotein
F:1-240/product: outer capsid protein VP8 #status predicted <VP8>
F:241-246/Region: cleavage processing #status predicted
F:247-775/Product: outer capsid protein VP5 #status predicted <VP5>
F:32,56,97,132,150,195,324,583,589,599/Binding site: carbohydrate (Asn) (covalent) #s

Query Match 49.4%; Score 42; DB 1; Length 775;
Best Local Similarity 43.8%; Pred. No. 36;
Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

OY 1 NMLVPMNPVLMNEFMN 16
DB 741 LNLIKSNPNVLRNFN 756

RESULT 6

VPXRK
outer layer protein VP3 - human rotavirus A (strain L27)
N:Alternate names: VP4 protein
C:Species: human rotavirus A
C:Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 01-Mar-1996
C:Accession: C36410
R:Taniguchi, K.; Urasawa, T.; Kobayashi, N.; Gorziglia, M.; Urasawa, S.
J. Virol. 64, 5640-5644, 1990
A:Title: Nucleotide sequence of VP4 and VP7 genes of human rotaviruses with subgroup
A:Reference number: A36410; MUID:91012813
A:Accession: C36410
A:Molecule type: genomic RNA
A:Residues: 1-775 <TRAN>
A:Cross-references: EMBL:M58292
C:Genetics:
A:Map position: segment 4
C:Superfamily: rotavirus outer layer protein VP3
C:Keywords: capsid protein; glycoprotein
F:1-240/product: outer capsid protein VP8 #status predicted <VP8>
F:241-246/Region: cleavage processing #status predicted
F:247-775/Product: outer capsid protein VP5 #status predicted <VP5>
F:32,56,97,132,150,195,324,583,589,599/Binding site: carbohydrate (Asn) (covalent) #s

Query Match 49.4%; Score 42; DB 1; Length 775;
Best Local Similarity 43.8%; Pred. No. 36;
Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

OY 1 NMLVPMNPVLMNEFMN 16
DB 741 LNLIKSNPNVLRNFN 756

RESULT 7

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 6, 2002, 11:02:28 ; Search time 121.65 seconds
(without alignments)
94.948 Million cell updates/sec

Title: US-09-897-042-17

Perfect score: 51
Sequence: 1 atgaactgtagcagatgaa.....tgaacgaatcagactaa 51

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	21.4	42.0	1912	1 US-08-270-013B-1	Sequence 1, Appli
2	21.4	42.0	1912	1 US-08-838-418-1	Sequence 1, Appli
3	21.2	41.6	805	3 US-08-961-083-139	Sequence 139, App
4	21.2	41.6	1945	1 US-08-724-194-1	Sequence 1, Appli
5	21.2	41.6	2678	1 US-08-724-194-2	Sequence 2, Appli
6	20.8	40.8	408	4 US-09-328-111-169	Sequence 169, App
7	20.8	40.8	3063	2 US-08-184-009-169	Sequence 169, App
8	20.8	40.8	3063	2 US-08-458-356-169	Sequence 169, App
9	20.8	40.8	3063	2 US-08-460-736-169	Sequence 169, App
10	20.8	40.8	3706	2 US-08-566-398-59	Sequence 59, Appl
11	20.8	40.8	3706	2 US-08-658-665-63	Sequence 63, Appl
12	20.8	40.8	3706	4 US-08-796-101-27	Sequence 27, Appl
13	20.8	40.8	3706	4 US-09-085-273-63	Sequence 63, Appl
14	20.6	40.4	176	2 US-08-687-080-82	Sequence 82, Appl
15	20.6	40.4	486	4 US-09-217-609A-28	Sequence 28, Appl
16	20.6	40.4	486	4 US-08-873-235B-28	Sequence 28, Appl
17	20.6	40.4	802	4 US-08-998-416-376	Sequence 376, App
18	20.6	40.4	1243	2 US-08-687-080-53	Sequence 53, Appl
19	20.6	40.4	5893	1 US-08-592-126-54	Sequence 54, Appl
20	20.6	40.4	5893	2 US-08-687-080-44	Sequence 44, Appl
21	20.4	40.0	420	4 US-09-214-095D-115	Sequence 115, App
22	20.4	40.0	423	1 US-08-470-179-120	Sequence 120, App
23	20.4	40.0	1854	1 US-08-249-420-1	Sequence 1, Appli
24	20.4	40.0	1854	2 US-08-737-663-1	Sequence 1, Appli
25	20.2	39.6	638	1 US-08-469-667-1	Sequence 1, Appli
26	20.2	39.6	638	5 PCT-US95-07289-1	Sequence 1, Appli
27	20.2	39.6	56516	2 US-08-996-306-1	Sequence 1, Appli

28	20.2	39.6	56516	4 US-09-338-907-1	Sequence 1, Appli
29	20.2	39.6	56520	1 US-09-338-907-179	Sequence 179, App
30	19.8	38.8	423	1 US-08-470-179-39	Sequence 39, Appl
31	19.8	38.8	423	1 US-08-470-179-45	Sequence 45, Appl
32	19.8	38.8	423	1 US-08-470-179-53	Sequence 53, Appl
33	19.8	38.8	899	1 US-07-820-154A-3	Sequence 3, Appli
34	19.8	38.8	899	2 US-08-097-954A-3	Sequence 3, Appli
35	19.8	38.8	899	3 US-08-480-640A-3	Sequence 3, Appli
36	19.8	38.8	899	3 US-08-295-802-3	Sequence 3, Appli
37	19.8	38.8	899	4 US-08-488-237A-3	Sequence 3, Appli
38	19.8	38.8	899	5 PCT-US93-0032A-3	Sequence 3, Appli
39	19.8	38.8	3628	3 US-08-480-640A-113	Sequence 113, App
40	19.8	38.8	3628	3 US-08-295-802-113	Sequence 113, App
41	19.8	38.8	3628	4 US-08-488-237A-113	Sequence 113, App
42	19.8	38.8	3942	3 US-08-480-640A-189	Sequence 189, App
43	19.8	38.8	3942	4 US-08-686-968C-189	Sequence 189, App
44	19.8	38.8	3942	4 US-08-488-237A-189	Sequence 189, App
45	19.8	38.8	5785	3 US-08-480-640A-221	Sequence 221, App

ALIGNMENTS

```

RESULT 1
US-08-270-013B-1
: Sequence 1, Application US/08270013B
: Patent No. 5686294
:
: GENERAL INFORMATION:
: APPLICANT: Sogabe et al.
: TITLE OF INVENTION: PROTEIN HAVING HEAT-RESISTANT MALATE
: NUMBER OF INVENTIONS: 2
: CORRESPONDENCE ADDRESSES:
: ADDRESSEE: Leydig, Volt & Mayer, Ltd.
: STREET: Two Prudential Plaza, Suite 4900
: CITY: Chicago
: STATE: Illinois
: COUNTRY: USA
: ZIP: 61601-6780
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/270,013B
: FILING DATE: 01-JUL-1994
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: JP 164701/1993
: FILING DATE: 02-JUL-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Green, Robert F.
: REGISTRATION NUMBER: 27555
: REFERENCE/DOCKET NUMBER: 62321
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (312) 616-5600
: TELEFAX: (312) 616-5700
: TELEX: (25)3533
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1912 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: ORIGINAL SOURCE:
: ORGANISM: Bacillus stearothermophilus
: STRAIN: ATCC12016
: US-08-270-013B-1
  
```

Query Match 42.0%; Score 21.4; DB 1; Length 1912;
Best Local Similarity 66.0%; Pred. No. 11;
Matches 31; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

OY 5 acttgtagcagatgaaccactagtaacgaacgaattcatgaactaa 51
||||| 11 11111 1 11111 11 1111 11
Db 966 ATTGCGTACGACGACCAAAAATTATGACGACGACGACGAA 1012

RESULT 2

US-08-838-418-1
; Sequence 1, Application US/08838418
; Patent No. 5744342
; GENERAL INFORMATION:
; APPLICANT: Sogabe et al.
; TITLE OF INVENTION: PROTEIN HAVING HEAT-RESISTANT MALATE
; TITLE OF INVENTION: DEHYDROGENASE ACTIVITY
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Leydig, Volt & Mayer, Ltd.
; STREET: Two Prudential Plaza, Suite 4900
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60601-6780
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/838.418
; FILING DATE: 17-MAR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/270.013
; FILING DATE: 01-JUL-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 164701/1993
; FILING DATE: 02-JUL-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Hoover, Allen E.
; REGISTRATION NUMBER: 37354
; REFERENCE/DOCKET NUMBER: 78339
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 616-5600
; TELEFAX: (312) 616-5700
; TELEX: (25)353
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1912 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Bacillus stearothermophilus
; STRAIN: ATCC12016
; US-08-838-418-1

Query Match 42.0%; Score 21.4; DB 1; Length 1912;
Best Local Similarity 66.0%; Pred. No. 11;
Matches 31; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

OY 5 acttgtagcagatgaaccactagtaacgaacgaattcatgaactaa 51
||||| 11 11111 1 11111 11 1111 11
Db 966 ATTGCGTACGACGACCAAAAATTATGACGACGACGACGAA 1012

RESULT 3
US-08-961-083-139/c
; Sequence 139, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 Inch, 1.4mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,083
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 139:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 805 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-961-083-139

Query Match 41.6%; Score 21.2; DB 3; Length 805;
Best Local Similarity 64.0%; Pred. No. 11;
Matches 32; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

OY 1 atgaactggtacgcagatgaaccactagtaacgaacgaattcatgaactaa 50
||||| 11 11111 11 1 11111 1 11111 1
Db 638 ATGAATATTGTTCTGATGATCCACCATTTGTAACATATCTCTGACCA 589

RESULT 4

US-08-724-194-1/c
; Sequence 1, Application US/08724194
; Patent No. 5824875
; GENERAL INFORMATION:
; APPLICANT: RANDU, RAJINDER S.
; TITLE OF INVENTION: ONE-AMINO CYCLOPROPANE-1-CARBOXYLATE
; TITLE OF INVENTION: SYNTHASE GENES FROM PELARCONIUM TO CONTROL ETHYLENE LEVELS
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SANTANGELO LAW OFFICES PC
; STREET: 315 WEST OAK STREET, STE 701
; CITY: FORT COLLINS
; STATE: CO
; COUNTRY: USA
; ZIP: 80521
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 6, 2002, 11:06:56 ; Search time 256.95 Seconds
(without alignments)
170.164 Million cell updates/sec

Title: US-09-897-042-17

Perfect score: 51
Sequence: 1 atgaacttgtagcagatgaa.....tgaacgaattcagactaa 51

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: N_Geneseq_1101: *
2: /SID52/gcgdata/geneseq/geneseqn/NA1980.DAT: *
3: /SID52/gcgdata/geneseq/geneseqn/NA1981.DAT: *
4: /SID52/gcgdata/geneseq/geneseqn/NA1982.DAT: *
5: /SID52/gcgdata/geneseq/geneseqn/NA1983.DAT: *
6: /SID52/gcgdata/geneseq/geneseqn/NA1984.DAT: *
7: /SID52/gcgdata/geneseq/geneseqn/NA1985.DAT: *
8: /SID52/gcgdata/geneseq/geneseqn/NA1986.DAT: *
9: /SID52/gcgdata/geneseq/geneseqn/NA1987.DAT: *
10: /SID52/gcgdata/geneseq/geneseqn/NA1988.DAT: *
11: /SID52/gcgdata/geneseq/geneseqn/NA1989.DAT: *
12: /SID52/gcgdata/geneseq/geneseqn/NA1990.DAT: *
13: /SID52/gcgdata/geneseq/geneseqn/NA1991.DAT: *
14: /SID52/gcgdata/geneseq/geneseqn/NA1992.DAT: *
15: /SID52/gcgdata/geneseq/geneseqn/NA1993.DAT: *
16: /SID52/gcgdata/geneseq/geneseqn/NA1994.DAT: *
17: /SID52/gcgdata/geneseq/geneseqn/NA1995.DAT: *
18: /SID52/gcgdata/geneseq/geneseqn/NA1996.DAT: *
19: /SID52/gcgdata/geneseq/geneseqn/NA1997.DAT: *
20: /SID52/gcgdata/geneseq/geneseqn/NA1998.DAT: *
21: /SID52/gcgdata/geneseq/geneseqn/NA1999.DAT: *
22: /SID52/gcgdata/geneseq/geneseqn/NA2000.DAT: *
23: /SID52/gcgdata/geneseq/geneseqn/NA2001.DAT: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	24.4	47.8	3246	22	AAH53972
2	24	47.1	630	20	AAK61539
3	24	47.1	116277	20	AAK20249
4	23.2	45.5	2244	12	AAQ11658
5	23.2	45.5	6706	22	AAK89648
6	23.2	45.5	6708	20	AAK86463
7	23.2	45.5	8897	20	AAK86464
8	23.2	45.5	8897	22	AAK89649
9	22.8	44.7	4191	19	AAV33692
10	22.6	44.3	7207	21	AAK21316
11	22.6	44.3	7207	21	AAK53194

12	22.6	44.3	15630	21	AAK21317
13	22.6	44.3	15630	21	AAK53195
14	22.2	43.5	4783	21	AAK76742
15	22	43.1	1830121	17	AAK42063
16	21.8	42.7	371	22	AAK67075
17	21.8	42.7	558	22	AAH10735
18	21.8	42.7	1599	21	AAK97797
19	21.8	42.7	2683	22	AAH17616
20	21.8	42.7	4115	21	AAK97430
21	21.6	42.4	245	20	AAK35521
22	21.6	42.4	300	20	AAK213737
23	21.6	42.4	300	20	AAK8348
24	21.6	42.4	551	22	AAH11430
25	21.6	42.4	753	20	AAK215437
26	21.6	42.4	1761	21	AAK45339
27	21.6	42.4	2522	22	AAH16195
28	21.4	42.0	171	22	AAK1744
29	21.4	42.0	171	22	AAK1745
30	21.4	42.0	510	21	AAK50790
31	21.4	42.0	1132	21	AAK38850
32	21.4	42.0	1912	17	AAK17715
33	21.4	42.0	2607	19	AAK19464
34	21.4	42.0	127197	22	AAK161370
35	21.4	42.0	335913	22	AAK161371
36	21.4	42.0	335913	22	AAK161372
37	21.2	41.6	322	20	AAK87277
38	21.2	41.6	805	19	AAK27393
39	21.2	41.6	1617	22	AAK53870
40	21.2	41.6	1934	18	AAK66246
41	21.2	41.6	1945	19	AAK30324
42	21.2	41.6	2678	22	AAK30325
43	21.2	41.6	3073	22	AAK54242
44	21.2	41.6	3723	22	AAK54274
45	21.2	41.6	4323	22	AAK58257

ALIGNMENTS

RESULT 1
ID AAK53972 standard; DNA; 3246 BP.
AC AAK53972:
DT 03-SEP-2001 (first entry)
DE S. epidermidis genomic polynucleotide sequence SEQ ID NO:3336.
XX
XX Staphylococcus epidermidis SRI strain; infection; diagnosis;
KW vaccination; endocarditis; ds.
XX
XX Staphylococcus epidermidis.
OS
XX
XX WO200134809-A2.
XX
XX 17-MAY-2001.
XX
XX 09-NOV-2000; 2000WO-US30782.
XX
XX 09-NOV-1999; 99US-0164258.
XX
XX (GLAXO) GLAXO GROUP LTD.
XX
XX kimberly WJ.
XX
XX WPI; 2001-316495/33.
XX
XX Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
XX useful for vaccinating against infections, e.g. endocarditis -
XX
XX Claim 8; Page 881-882; 2188pp; English.
XX

CC AAH5204 to AAH53970 represent nucleic acids (I) encoding polypeptides
CC (II), given in AAG81454 to AAG83120, from *Staphylococcus epidermidis*.
CC (I) and (II) can have antibacterial activity and therefore can be used
CC in vaccination. The nucleic acids (I) may be used to produce the
CC S. epidermidis polypeptides (II) via the production of vectors
CC containing them which are used to produce hosts cells which express the
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
CC used to vaccinate subjects and to raise antibodies against the bacteria.
CC The polypeptides may also be used to assay for other inhibitors of their
CC activity and therefore identify compounds that may be used for the
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
CC AAH55090 represent specifically claimed S. epidermidis genomic DNA
CC polynucleotide sequences from the present invention. AAH5091 to
CC AAH5098 represent oligonucleotide sequences and primers which are used
CC in the exemplification of the present invention.
CC N.B. The present invention specifically claims all the polynucleotide
CC sequences given in the sequence listing of the present specification,
CC however the sequence listing only goes up to SEQ ID NO:4454 so even
CC though sequences are given in the disclosure for SEQ ID NO:4455 to 4472,
CC no sequences are present for SEQ ID NO:4455 to 4464.
XX
XX
SO Sequence 3246 BP; 1091 A; 533 C; 536 G; 1086 T; 0 other;

Query Match 47.8%; Score 24.4; DB 22; Length 3246;
Best Local Similarity 68.0%; Pred. No. 4.3;
Matches 34; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Oy 1 atgaactgtgacagatgaccacactagtaatgaacgaattcagaacta 50
||| ||||| | |||| | ||||| ||||| ||| ||
Db 871 atgcttgcgaatgaagaagacaacatttaaaagaacgaattcaagaacta 920

RESULT 2

AAH61539
ID AAH61539 standard; DNA; 630 BP.

XX
AC AAX61539;

XX
DT 19-JUL-1999 (first entry)

XX
DE B. burgdorferi antigenic protein coding sequence, f516.nt.

XX
KW Antigenic protein; vaccine; Lyme disease; infection; detection; ss.

XX
OS Borrelia burgdorferi.

XX
PN WO9859071-A1.

XX
PD 30-DEC-1998.

XX
PF 18-JUN-1998; 98WO-US12718.

XX
PR 03-SEP-1997; 97US-0057483.

XX
PR 20-JUN-1997; 97US-0050359.

XX
PR 22-JUL-1997; 97US-0053344.

XX
PR 22-JUL-1997; 97US-0053377.

XX
PA (HUMA-) HUMAN GENOME SCI INC.

XX
PI (MEDI-) MEDIMUNE INC.

XX
PI Choi GH, Erwin AL, Hanson MS, Lathigra R;

XX
DR WPI: 1999-189980/16.

XX
DR P-PSDB; AAY1842.

XX
PT New isolated Borrelia burgdorferi nucleic acids - used to develop

XX
PT products for the diagnosis, prevention and treatment of diseases

XX
PT caused by Borrelia, particularly Lyme disease

XX
PS Claim 1; Page 88; 275pp; English.

XX
CC This sequence encodes a Borrelia burgdorferi (Bb) protein of the

CC invention, which is suitable for use in a vaccine. The Bb polypeptides
CC can be used in vaccines for eliciting protective antibodies to members of
CC the Borrelia genus, particularly for the use against Lyme disease in
CC humans and animals. They can be used for preventing or attenuating an
CC infection caused by a member of the Borrelia genus. The products can also
CC be used for detection of members of the Borrelia genus.
XX
XX
SO Sequence 630 BP; 233 A; 92 C; 81 G; 224 T; 0 other;

Query Match 47.1%; Score 24; DB 20; Length 630;
Best Local Similarity 68.8%; Pred. No. 4.6;
Matches 33; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Oy 1 atgaactgtgacagatgaccacactagtaatgaacgaattcagaac 48
||| ||||| | |||| | ||||| ||||| ||| |||
Db 59 atgcttgcgaatgaagaagagcaactactaatgaataaaatgacaaac 106

RESULT 3

AAX20249
ID AAX20249 standard; DNA; 116277 BP.

XX
AC AAX20249;

XX
DT 04-MAY-1999 (first entry)

XX
DE Borrelia burgdorferi polynucleotide sequence #2.

XX
KW Borrelia burgdorferi; spirochete; bacterium; pathogen; Lyme disease;

XX
KW epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;

XX
KW infection; diagnosis; characterisation; detection; ds.

XX
OS Borrelia burgdorferi.

XX
PN WO9858943-A1.

XX
PD 30-DEC-1998.

XX
PF 18-JUN-1998; 98WO-US12764.

XX
PR 03-SEP-1997; 97US-0057483.

XX
PR 20-JUN-1997; 97US-0050359.

XX
PR 22-JUL-1997; 97US-0053344.

XX
PR 22-JUL-1997; 97US-0053377.

XX
PA (HUMA-) HUMAN GENOME SCI INC.

XX
PI (MEDI-) MEDIMUNE INC.

XX
PI Clayton R, Dougherty BA, Fraser C, Lathigra R, Smith HO;

XX
PI White OR;

XX
DR WPI: 1999-081217/07.

XX
PT New isolated Borrelia burgdorferi nucleic acids - used to develop

XX
PT products for the detection, diagnosis, characterisation, prevention

XX
PT and therapy of infections, particularly Lyme disease

XX
PS Claim 1; Page 672-737; 1128pp; English.

XX
CC AAX20248 to AAX20402 represent polynucleotide sequences isolated from

XX
CC Borrelia burgdorferi (Bb). Products derived from Bb can be used for

XX
CC the detection, diagnosis, characterisation, prevention and therapy of

XX
CC Bb infections, e.g. Lyme disease. They can also be used for the

XX
CC production of biosynthetic products, e.g. enzymes. Borrelia belongs

XX
CC to a family of motile, spiral-shaped bacteria called Spirochetes.

XX
CC Spirochetes are pathogenic in humans and Borrelia causes epidemic and

XX
CC endemic relapsing fever, and Lyme borreliosis, more commonly known as

XX
CC Lyme disease.

XX
SO Sequence 116277 BP; 42656 A; 19868 C; 14490 G; 39250 T; 13 other;

GenCore version 4.5
Copyright (c) 1993 - 2000 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 6, 2002, 10:33:30 ; Search time 2642.04 Seconds
(without alignments)
207.429 Million cell updates/sec

Title: US-09-897-042-17

Perfect score: 51
Sequence: 1 atgaactgtgaccatgaa.....tgaacgaatcatgaactaa 51

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_hic:*
10: gb_est1:*
11: gb_est2:*
12: gb_hic:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rtd:*
20: em_gss_vit:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	25.8	50.6	511	11	BG041695 sv37g01.Y
2	25	49.0	607	11	C84717 C84717 Dict
3	24.8	48.6	465	10	AM641204 CM04604.V
4	24.6	48.2	883	13	CNS07091 AL440171 T3 end of
5	24.4	47.8	348	10	AV548944 AV548944
6	24.4	47.8	506	10	AV543742 AV543742
7	24.4	47.8	529	10	AV544086 AV544086
8	24.4	47.8	538	13	AO968417 LERJC35TR
9	24.4	47.8	655	10	AV545019 AV545019
10	24.4	47.8	662	13	AO968416 LERJC35TF
11	24.2	47.5	682	13	AO795238 nbxb0055N
12	24.2	47.5	761	13	AO326076 nbxb0024J

C 13	24.2	47.5	776	13	AO576890 nbxb0089P
C 14	24	47.1	843	13	CNS03BRH
C 15	23.8	46.7	392	10	AM371950 RC4-BP031
C 16	23.8	46.7	419	13	AO563015 HS-5301_A
C 17	23.8	46.7	1070	13	CNS060PEX T3 end of
C 18	23.6	46.3	365	11	BF415704 BF415704
C 19	23.6	46.3	528	11	BF416556 BF416556
C 20	23.6	46.3	535	13	AZ467538 LM0279804
C 21	23.6	46.3	570	13	AZ371615 LM0123D12
C 22	23.6	46.3	575	13	AZ371859 LM0133P22
C 23	23.6	46.3	793	10	BE130386 L48-560T3
C 24	23.6	46.3	964	11	BE387522 602412333
C 25	23.6	46.3	1187	11	BF796700 602258321
C 26	23.4	45.9	363	13	AO033770 HS-2237_A
C 27	23.4	45.9	471	11	N91806 N91806
C 28	23.4	45.9	496	11	BG318404 NXPV_013-
C 29	23.4	45.9	537	11	BI202938 NXPV_092-
C 30	23.4	45.9	557	11	BG318869 NXPV_020-
C 31	23.4	45.9	595	10	AA495617 c428_Zhou
C 32	23.4	45.9	967	13	CNS06PL6
C 33	23.2	45.5	272	11	BG124644 CC-estf1cl
C 34	23.2	45.5	404	11	BG629894 CC-estf1cl
C 35	23.2	45.5	522	11	BF112629 EST440219
C 36	23	45.1	265	10	BS592896 BS592896
C 37	23	45.1	360	11	BI424347 BS631903
C 38	23	45.1	768	13	AO420780 RPTC1-11-2
C 39	23	45.1	844	13	AZ534572 EMTCT77TR
C 40	22.8	44.7	255	10	AI071750 UI-R-C2-n
C 41	22.8	44.7	307	11	N59695 YV55905_r1
C 42	22.8	44.7	459	10	AV542740 AV542740
C 43	22.8	44.7	467	10	BE587941 WHE0665_F
C 44	22.8	44.7	562	11	BG135904 EST476346
C 45	22.8	44.7	567	13	AZ711369 RPTC1-24-1

ALIGNMENTS

RESULT 1	BG041695	511 bp	mRNA	EST	31-JUL-2001
LOCUS	sv37g01.Y1	Gm-cl057	glycine max	CDNA clone	GENOME SYSTEMS CLONE ID:
DEFINITION	Gm-cl057-1849 5'	similar to	TR:Q9SH32 Q9SH32 F2K11.16	1	mRNA
ACCESSION	BG041695	GI:12487985			
VERSION	BG041695				
KEYWORDS	EST				
SOURCE	soybean				
ORGANISM	Glycine max				
REFERENCE	1 (bases 1 to 511)				
AUTHORS	Shoemaker,R., Keim,P., Vodka,L., Erpellding,J., Coryell,V., Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Stepien,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterson,R. and Wilson,R.				
TITLE	Public Soybean EST Project				
JOURNAL	Unpublished (1999)				
COMMENT	Contact: Shoemaker R/Public Soybean EST Project Public Soybean EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or				

BASE COUNT ,	215 a	103 c	67 g	222 t
--------------	-------	-------	------	-------

COMMENT

Contact: Perry J. Blackshear
Office of Clinical Research and Laboratory of Signal Transduction
National Institute of Environmental Health Sciences
A2-05 NIHMS, 101 Alexander Drive, Research Triangle Park, NC 27709,
USA
Tel: 919 541-4899
Fax: 919 541-4571
Email: black009@niehs.nih.gov
Clone is available through Research Genetics, Inc., 2130 Memorial
Parkway, Huntsville, AL 35901
phone 800-533-4363 ext.cdna, fax 256-536-9016 att:cdna, email
cna@resgen.com
DNA Sequencing and analyses performed by National Institutes of
Health Intramural Sequencing Center (NISC).

PCR primers
FORWARD: TGTTAAACGACGGCCAGT
BACKWARD: CAGGAACAGCTATGAC
Plate: 0103 row: E column: 04
Seq primer: T7 primer
Location/Qualifiers
1..465

FEATURES

Source

.

/nucleo-/vector: p7733-Pac; Site1: SCORI; Site2: Noli; POLYA-selected mRNA was prepared from unfertilized *Xenopus* laevis eggs. The library was constructed in the vector p7733-Pac as described in Bonaldo, M. F., Lennon, G. and Soares, M. B. 'Normalization and subtraction: two approaches to facilitate gene discovery', *Genome Research* 6:791-806, 1996. The first strand synthesis used a Noli-d18 primer; double stranded cDNAs were ligated to Noli adapters, digested with Noli, and directionally

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 6, 2002, 11:00:38 ; Search time 1586.23 Seconds

(without alignments)
530.412 Million cell updates/sec

Title: US-09-897-042-17

Perfect score: 51
Sequence: 1 atgaacttgtagcagatgaa.....tgaacgaatcatgaactaa 51

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GeneBdb1:*
1: gb_ba:*
2: gb_mt9:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_om:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vl:*
30: em_higo_hum:*
31: em_higo_inv:*
32: em_higo_rod:*
33: em_hig_hum:*
34: em_hig_inv:*
35: em_hig_rod:*
36: em_hig_other:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	26.4	51.8	8601	3	AF274754	AF274754 Brugia ma
2	24.8	48.6	168497	2	AC092410	AC092410 Bos tauru
3	24.6	48.2	4536	14	F1PFP	D32044 Feline infe
C 4	24.6	48.2	300950	1	AP001516	AP001516 Bacillus
5	24.4	47.8	3246	1	AF269294	AF269294 Staphyloc
6	24.4	47.8	3246	6	AX144614	AX144614 Sequence
C 7	24.4	47.8	57198	2	AC020115	AC020115 Drosophila
8	24.4	47.8	118507	8	ATF7J8	AL137189 Arabidops
C 9	24.4	47.8	169090	3	AC007300	AC007300 Drosophila
10	24.4	47.8	260791	3	AE003629	AE003629 Drosophila
11	24.4	47.1	18604	1	AE001146	AE001146 Borrelia
12	24.4	47.1	35050	2	AC006099	AC006099 Homo sapi
13	24.4	47.1	67249	2	AC018611	AC018611 Homo sapi
C 14	24.4	47.1	67249	2	AC018611	AC018611 Homo sapi
15	24.4	47.1	163729	2	AC013252	AC013252 Homo sapi
C 16	24.4	47.1	167501	2	AL391421	AL391421 Homo sapi
17	24.4	47.1	171505	9	CNS01DTH	AL133719 Human chr
C 18	24.4	47.1	180506	2	AC069306	AC069306 Homo sapi
19	24.4	47.1	219476	9	AC068945	AC068945 Homo sapi
20	23.6	46.3	6380	8	PETN1AA	U13563 Petunia hyb
21	23.6	46.3	6380	8	PETN1TRPD	U13563 Petunia hyb
C 22	23.6	46.3	134534	2	AL512352	AL512352 Homo sapi
C 23	23.6	46.3	152027	2	AC027386	AC027386 Homo sapi
24	23.6	46.3	157657	9	AC022841	AC022841 Homo sapi
C 25	23.6	46.3	174494	2	AC068744	AC068744 Homo sapi
26	23.6	46.3	190056	2	AP002430	AP002430 Homo sapi
C 27	23.6	46.3	190642	2	AC062013	AC062013 Homo sapi
28	23.4	45.9	3413	5	CHA249579	AL249579 Chionodra
C 29	23.4	45.9	166036	2	AC041013	AC041013 Homo sapi
30	23.4	45.9	179702	2	AC090058	AC090058 Homo sapi
31	23.4	45.9	205363	2	AC078998	AC078998 Mus muscu
32	23.2	45.5	2239	8	SLU60267	U60267 Solanum lyc
C 33	23.2	45.5	8897	14	TSWRLPOLM	D10066 Tomato spot
34	23.2	45.5	182056	9	AC073533	AC073533 Homo sapi
C 35	23.2	45.1	23830	8	SPAC27F1	269368 S. pombe chr
C 36	23.2	45.1	40029	3	CEP25C8	281512 Caenorhabdi
C 37	23.2	45.1	43160	2	CEP14H10	292763 Caenorhabdi
38	23.2	45.1	62844	2	AC023982	AC023982 Homo sapi
C 39	23.2	45.1	127590	9	AC002554	AC002554 Human chr
C 40	23.2	45.1	152567	33	AC011834	AC011834 Homo sapi
41	23.2	45.1	161737	9	AC090512	AC090512 Homo sapi
42	23.2	45.1	174194	2	AL512329	AL512329 Homo sapi
43	22.8	44.7	2600	4	EKU97342	U97342 Echinipera
44	22.8	44.7	4191	1	LLPFLMG13	AJ000325 Lactococc
45	22.8	44.7	4481	3	DCU94491	U94491 Dictyostell

ALIGNMENTS

RESULT	1	AF274754/c	LOCUS	AF274754	8601 bp	DNA	INV	23-JAN-2001
DEFINITION				Brugia malayi chitin synthase (chsl) gene, complete cds.				
ACCESSION				AF274754				
VERSION				AF274754.1	GI:12382255			
KEYWORDS								
SOURCE				Brugia malayi				
ORGANISM				Brugia malayi				
				Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;				
				Onchocercidae; Brugia				
REFERENCE				1 (bases 1 to 8601)				
AUTHORS				Harris, M.T., Lai, K., Arnold, K., Martinez, H.F., Specht, C.A. and				
				Fuhrman, J.A.				
TITLE				Chitin synthase in the filarial parasite, Brugia malayi				
JOURNAL				Mol. Biochem. Parasitol. 111 (2), 351-362 (2000)				
MEDLINE				21036600				
REFERENCE				2 (bases 1 to 8601)				
AUTHORS				Fuhrman, J.A., Arnold, K., Lai, K. and Martinez, H.F.				
TITLE				Direct Submission				

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 6, 2002, 09:15:14 ; Search time 71.9 Seconds

(without alignments)
14.710 Million cell updates/sec

Title: US-09-897-042-16

Perfect score: 262

Sequence: 1 MNLVPMNPLVMNGFCRYPH.....RLTKPRRLSWLLPLLSNN 47

Scoring table: BLOSUM62

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Post-processing: Minimum Match 08

Maximum Match 100%
Listing first 50 summaries

Database :

1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCRTUS.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	195.5	74.6	314	1	US-08-040-753-2
2	59	22.5	110	1	US-08-081-539-69
3	59	22.5	110	1	US-08-466-647-69
4	53.5	20.4	124	1	US-08-240-049B-20
5	53.5	20.4	124	4	US-08-542-634-22
6	53.5	20.4	124	4	US-08-477-292-22
7	53.5	20.4	124	5	PCR-US95-13703-22
8	52.5	20.0	346	1	US-07-895-300A-1
9	52.5	20.0	346	1	US-08-458-367-1
10	52.5	20.0	346	5	PCR-US93-05136-1
11	52	19.8	923	3	US-08-936-135-6
12	51.5	19.7	104	2	US-08-341-843B-35
13	51.5	19.7	104	2	US-08-427-497E-40
14	51.5	19.7	894	4	US-08-599-455B-2
15	51.5	19.7	894	4	US-09-069-781B-2
16	51.5	19.7	953	2	US-08-500-857A-2
17	51.5	19.7	1162	2	US-08-599-455B-43
18	51.5	19.7	1162	4	US-09-069-781B-43
19	50	19.1	454	3	US-08-929-329-8
20	49.5	18.9	142	2	US-08-392-546C-1
21	49.5	18.9	896	2	US-08-640-389A-12
22	48.5	18.5	405	1	US-08-351-473B-2
23	48.5	18.5	895	4	US-08-827-962-19
24	48.5	18.5	895	4	US-08-827-962-21
25	48.5	18.5	1162	4	US-08-827-962-15
26	48.5	18.5	1162	4	US-08-827-962-20
27	48.5	18.5	1162	4	US-08-803-346-1

28	48	18.3	379	3	US-08-945-056-2	Sequence 2, Appl1
29	48	18.3	379	1	US-08-164-614A-8	Sequence 8, Appl1
30	48	18.3	379	2	US-08-456-489B-8	Sequence 8, Appl1
31	48	18.3	468	1	US-08-164-614A-7	Sequence 7, Appl1
32	48	18.3	468	2	US-08-456-489B-7	Sequence 7, Appl1
33	47.5	18.1	120	1	US-08-497-312-26	Sequence 26, Appl1
34	47.5	18.1	120	1	US-08-497-312-28	Sequence 26, Appl1
35	47.5	18.1	373	3	US-08-911-321-9	Sequence 28, Appl1
36	47.5	18.1	373	5	PCR-US95-13975-73	Sequence 73, Appl1
37	47.5	18.1	374	1	US-08-220-151-21	Sequence 21, Appl1
38	47.5	18.1	374	1	US-08-413-118-21	Sequence 21, Appl1
39	47.5	18.1	374	3	US-08-473-446-21	Sequence 21, Appl1
40	47	17.9	339	3	US-08-968-563-14	Sequence 14, Appl1
41	47	17.9	339	4	US-08-969-683A-14	Sequence 14, Appl1
42	47	17.9	375	1	US-08-468-847B-13	Sequence 13, Appl1
43	47	17.9	432	4	US-08-702-665A-3	Sequence 3, Appl1
44	47	17.9	441	4	US-09-151-102-4	Sequence 4, Appl1
45	46.5	17.7	550	1	US-08-484-493-2	Sequence 2, Appl1
46	46.5	17.7	550	1	US-08-484-494-2	Sequence 2, Appl1
47	46.5	17.7	550	2	US-08-345-212-2	Sequence 2, Appl1
48	46.5	17.7	550	4	US-09-249-003-2	Sequence 2, Appl1
49	46.5	17.7	591	1	US-08-145-995A-21	Sequence 21, Appl1
50	46.5	17.7	591	2	US-08-451-747-21	Sequence 21, Appl1

ALIGNMENTS

RESULT 1
US-08-040-753-2
; Sequence 2, Application US/08040753
; Patent No. 5464745
; GENERAL INFORMATION:
; APPLICANT: Mierendorf, Robert
; APPLICANT: Garber, Richard
; APPLICANT: No. 5464745y, Robert
; APPLICANT: Hammer, Beth
; TITLE OF INVENTION: Protein Ligand Binding
; TITLE OF INVENTION: Region Mapping System
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles and Brady
; STREET: 1 South Pinckney St., Box 2113
; CITY: Madison
; STATE: WI
; COUNTRY: USA
; ZIP: 53701-2113
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/040,753
; FILING DATE: 19930331
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J
; REGISTRATION NUMBER: 27386
; REFERENCE/DOCKET NUMBER: 70-399-9001-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 608-251-5000
; TELEFAX: 608-251-9166
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 314 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-040-753-2

Query Match 74.6%; Score 195.5; DB 1; Length 314;
Best Local Similarity 76.5%; Pred. No. 6,4e-18;
Matches 39; Conservative 3; Mismatches 4; Indels 5; Gaps 2;

QY 2 NLVPMPLVWNGFCRYPSPHMRPLEQIRLTKPERRLSWLLPPLSN 47
DB 264 SLVPSDPLVTAASVLECRYPSPHMRPLEQIRLTKPERRLSWLLPPLSN 314

RESULT 2

US-08-081-539-69
Sequence 69, Application US/08081539
Patent No. 5501962
GENERAL INFORMATION:
APPLICANT: Braford-Goldberg, Sarah R.
APPLICANT: Easton, Alan M.
APPLICANT: Klein, Barbara K.
APPLICANT: McKeam, John P.
APPLICANT: Olin, Peter O.
TITLE OF INVENTION: Interleukin-3 (IL-3) Human/Murine
TITLE OF INVENTION: Chimeric Hybrid Polypeptides
NUMBER OF SEQUENCES: 121
CORRESPONDENCE ADDRESS:
ADDRESSEE: Mary Jo Kanady, G. D. Searle & Co., Corporate
ADDRESSEE: Patent Dept.
STREET: P. O. Box 5110
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60680
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/081,539
FILING DATE: 19930621
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Kanady, Mary J.
REGISTRATION NUMBER: 28623
REFERENCE/DOCKET NUMBER: 2724
TELECOMMUNICATION INFORMATION:
TELEPHONE: (708)470-6501
TELEFAX: (708)470-6881
INFORMATION FOR SEQ ID NO: 69:
SEQUENCE CHARACTERISTICS:
LENGTH: 110 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-081-539-69

Query Match 22.5%; Score 59; DB 1; Length 110;
Best Local Similarity 33.3%; Pred. No. 1.1;
Matches 15; Conservative 8; Mismatches 16; Indels 6; Gaps 1;

QY 2 NLVPMPLVWNGFCRYPSPHMRPLEQIRLTKPERRLSWLLPPLSN 46
DB 67 NLPLCLPLATAAPTRRHPIHKKDDEFR-----RKLTFYLTLEN 105

RESULT 3

US-08-466-647-69
Sequence 69, Application US/08466647
Patent No. 5543141
GENERAL INFORMATION:
APPLICANT: Braford-Goldberg, Sarah R.
APPLICANT: Easton, Alan M.
APPLICANT: Klein, Barbara K.
APPLICANT: McKeam, John P.

APPLICANT: Olin, Peter O.
TITLE OF INVENTION: Interleukin-3 (IL-3) Human/Murine
TITLE OF INVENTION: Chimeric Hybrid Polypeptides
NUMBER OF SEQUENCES: 121
CORRESPONDENCE ADDRESS:
ADDRESSEE: Mary Jo Kanady, G. D. Searle & Co., Corporate
ADDRESSEE: Patent Dept.
STREET: P. O. Box 5110
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60680

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/466,647
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/081,539
FILING DATE: 21-JUN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Kanady, Mary J.
REGISTRATION NUMBER: 28623
REFERENCE/DOCKET NUMBER: 2724
TELECOMMUNICATION INFORMATION:
TELEPHONE: (708)470-6501
TELEFAX: (708)470-6881
INFORMATION FOR SEQ ID NO: 69:
SEQUENCE CHARACTERISTICS:
LENGTH: 110 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-466-647-69

Query Match 22.5%; Score 59; DB 1; Length 110;
Best Local Similarity 33.3%; Pred. No. 1.1;
Matches 15; Conservative 8; Mismatches 16; Indels 6; Gaps 1;

QY 2 NLVPMPLVWNGFCRYPSPHMRPLEQIRLTKPERRLSWLLPPLSN 46
DB 67 NLPLCLPLATAAPTRRHPIHKKDDEFR-----RKLTFYLTLEN 105

RESULT 4

US-08-240-049B-20
Sequence 20, Application US/08240049B
Patent No. 5685239
GENERAL INFORMATION:
APPLICANT: Reyes, Gregory R.
APPLICANT: Tam, Albert W.
APPLICANT: Yarbough, Patricia O.
TITLE OF INVENTION: Hepatitis E Virus Peptides and Methods
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genelabs Technologies, Inc.
STREET: 505 Penobscot Drive
CITY: Redwood City
STATE: CA
COUNTRY: USA
ZIP: 94063
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/240,049B

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 6, 2002, 09:18:42 ; Search time 103.34 Seconds

(without alignments)
33.689 Million cell updates/sec

Title: US-09-897-042-16

Perfect score: 262

Sequence: 1 MNLVPMNPLVMNGFCRYPH.....RLTKPERRLSWLPLPSNN 47

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 segs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 50 summaries

Database : A.Geneseq.1101.*

1: /SID22/gcgdata/geneseq/geneseq/AA1980.DAT.*

2: /SID22/gcgdata/geneseq/geneseq/AA1981.DAT.*

3: /SID22/gcgdata/geneseq/geneseq/AA1982.DAT.*

4: /SID22/gcgdata/geneseq/geneseq/AA1983.DAT.*

5: /SID22/gcgdata/geneseq/geneseq/AA1984.DAT.*

6: /SID22/gcgdata/geneseq/geneseq/AA1985.DAT.*

7: /SID22/gcgdata/geneseq/geneseq/AA1986.DAT.*

8: /SID22/gcgdata/geneseq/geneseq/AA1987.DAT.*

9: /SID22/gcgdata/geneseq/geneseq/AA1988.DAT.*

10: /SID22/gcgdata/geneseq/geneseq/AA1989.DAT.*

11: /SID22/gcgdata/geneseq/geneseq/AA1990.DAT.*

12: /SID22/gcgdata/geneseq/geneseq/AA1991.DAT.*

13: /SID22/gcgdata/geneseq/geneseq/AA1992.DAT.*

14: /SID22/gcgdata/geneseq/geneseq/AA1993.DAT.*

15: /SID22/gcgdata/geneseq/geneseq/AA1994.DAT.*

16: /SID22/gcgdata/geneseq/geneseq/AA1995.DAT.*

17: /SID22/gcgdata/geneseq/geneseq/AA1996.DAT.*

18: /SID22/gcgdata/geneseq/geneseq/AA1997.DAT.*

19: /SID22/gcgdata/geneseq/geneseq/AA1998.DAT.*

20: /SID22/gcgdata/geneseq/geneseq/AA1999.DAT.*

21: /SID22/gcgdata/geneseq/geneseq/AA2000.DAT.*

22: /SID22/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	195.5	74.6	314	AA87027	T7 gene 10 leader
2	84	32.1	654	AA83378	Chlamydia trachoma
3	80.5	30.7	691	AA83371	Chlamydia trachoma
4	80	30.5	683	AA83382	Chlamydia trachoma
5	79	30.2	646	AA83372	Chlamydia trachoma
6	77.5	29.6	715	AA83373	Chlamydia trachoma
7	77	29.4	715	AA83375	Chlamydia trachoma
8	75	28.6	583	AA83381	Chlamydia trachoma
9	74.5	28.4	518	AA83376	Chlamydia trachoma
10	74	28.2	585	AA83377	Chlamydia trachoma
11	74	28.2	700	AA83379	Chlamydia trachoma

12	73	27.9	487	22	AA83380	Chlamydia trachoma
13	73	27.9	619	22	AA83370	Chlamydia trachoma
14	73	27.9	631	22	AA83374	Chlamydia trachoma
15	70	26.7	666	20	AA82471	Human C-term mychl
16	70	26.7	666	22	AA859203	C-terminal mychl
17	70	26.7	669	20	AA82470	Human C-term V5 his
18	70	26.7	669	22	AA859202	C-terminal V5 his
19	59	22.5	110	16	AA869757	PMON5968 (Met-15-1
20	59	22.5	132	22	AA840409	Human polypeptide
21	57	21.8	442	22	AA841638	Human polypeptide
22	56	21.4	281	22	AA848050	Signal transductio
23	56	21.4	391	22	AA87488	B thuringiensis 44
24	55.5	21.2	131	19	AA852130	Putative insectici
25	54.5	20.8	553	20	AA838796	Neisseria meningit
26	54.5	20.8	558	20	AA838797	Neisseria meningit
27	53.5	20.4	123	14	AA838786	HEV ORF3 protein.
28	53.5	20.4	123	14	AA839307	Mexico strain HEV
29	53.5	20.4	123	17	AA866098	Hepatitis E virus
30	53.5	20.4	123	20	AA893398	Human HEV ORF 3 pr
31	53.5	20.4	124	18	AA835829	Hepatitis E virus
32	53	20.2	132	9	AA80381	Interleukin-3 vari
33	53	20.2	133	9	AA80383	Interleukin-3 vari
34	53	20.2	155	21	AA811515	SEN virus protein
35	53	20.2	753	21	AA811525	SEN virus protein
36	53	20.2	1215	21	AA815791	Superheat-resistan
37	53	20.2	1215	22	AA84784	Pyrococcus heat re
38	52.5	20.0	51	22	AA817391	Peptide #3825 enco
39	52.5	20.0	51	22	AA829911	Peptide #3948 enco
40	52.5	20.0	51	22	AA805079	Peptide #3761 enco
41	52.5	20.0	260	17	AA814408	Human DNase I vari
42	52.5	20.0	260	17	AA814409	Human DNase I vari
43	52.5	20.0	260	17	AA814410	Human DNase I vari
44	52.5	20.0	260	17	AA814411	Human DNase I vari
45	52.5	20.0	260	17	AA814412	Human DNase I vari
46	52.5	20.0	260	17	AA814413	Human DNase I vari
47	52.5	20.0	260	17	AA814414	Human DNase I vari
48	52.5	20.0	260	17	AA814415	Human DNase I vari
49	52.5	20.0	260	17	AA814416	Human DNase I vari
50	52.5	20.0	260	17	AA814417	Human DNase I vari

ALIGNMENTS

RESULT 1	
AA87027	standard; Protein; 314 AA.
AC	AA87027;
XX	
DT	18-MAR-1996 (first entry)
XX	
DE	T7 gene 10 leader sequence product.
XX	
KW	Plasmid pTOPE-1b(+); vector; ligand binding domain; epitope mapping;
KM	antigen; Escherichia coli.
XX	
OS	Bacteriophage T7.
XX	
PN	US5464745-A.
XX	
PD	07-NOV-1995.
XX	
PF	31-MAR-1993; 93US-0040753.
XX	
PR	31-MAR-1993; 93US-0040753.
XX	
PA	(NOVA-) NOVAGEN INC.
XX	
PI	Garber R, Hammer B, Mierendorf R, Novy R;
XX	WPI; 1995-382610/50.
DR	N-PDB; AAT07310.

PT	Mapping ligand binding domains, esp. epitope(s), of proteins - by expressing peptide(s) encoded by random gene fragments and testing for ligand binding
XX	
PS	Disclosure; Column 15-18; 12pp; English.
CC	
CC	The phage T7 gene 10 leader sequence product (AAR67027) is encoded by prokaryotic expression vector pPROE-Tb(+)(AA07310). A fusion of the leader sequence and a putative ligand-binding domain (LBD), esp., antigenic region, of a protein is obtd. following insertion of putative LBD-encoding DNA into the vector. The fusion protein accumulates as inclusion bodies in Escherichia coli host cells and can be screened for its ability to bind a ligand.
CC	
XX	
SQ	Sequence 314 AA;
OY	Query Match 74.6%; Score 195.5; DB 16; Length 314; Best Local Similarity 76.5%; Pred. No. 4.9e-17; Matches 39; Conservative 3; Mismatches 4; Indels 5; Gaps 2
D6	2 NLVP--NNPLVNMNG---FCRRYPSHMRPLEQRILTKRPERRLSMLLPISNN 47 :: :: 264 slvpsadplvtasvleferypshwrpleqriltkperkswlllpisnn 314
RESULT 2	
ID	AAG83278 standard; Protein; 654 AA.
AC	AAG83278;
XX	
DT	05-SEP-2001 (first entry)
XX	
DE	Chlamydia trachomatis PmpB(3) fusion protein.
XX	
KW	Chlamydia; vaccine; infection; fusion protein; antigen; pelvic inflammatory disease; trachoma; atherosclerosis; heart disease; acute respiratory tract infection; CapI; CF529; OMCB;
KW	polymorphic membrane protein; pmp; thiol specific antioxidant; TSA.
XX	
OS	Chlamydia trachomatis.
XX	
PX	MO200140474-A2.
PD	07-JUN-2001.
XX	
PE	04-DEC-2000; 2000WO-US32919.
XX	
PR	03-DEC-1999; 99US-0454684. 19-APR-2000; 2000US-0556877. 20-JUN-2000; 2000US-0598419.
XX	
PA	(CORI-) CORIXA CORP.
PI	Probst P, Bhatia A, Skelky YAW, Fling SP, Scholler J;
DR	WPI; 2001-374831/39.
XX	
PT	Chlamydia polypeptides and fusion proteins useful for preventing pelvic inflammatory disease, trachoma, acute respiratory tract infections, atherosclerosis and heart disease -
XX	
PS	Claim 70; Page 284-285; 295pp; English.
XX	
CC	The present sequence is provided in a specification relating to compounds and methods for the treatment and diagnosis of chlamydial infection. The compounds provided include polypeptides and fusion proteins comprising immunogenic portions of Chlamydia antigens and DNA sequences encoding such polypeptides. They are useful for vaccinating against chlamydial infection, which causes pelvic inflammatory disease, trachoma, acute respiratory tract infections,
CC	

```

CC      atherosclerosis and heart disease.
XX
SQ      Sequence : 654 AA;

Query Match      32.1%; Score 84; DB 22; Length 654;
Best Local Similarity 75.0%; Pred. No. 0.018;
Matches 12; Conservative 3; Mismatches 1; Indels 0; Gaps 0.

QY      14 FCRYPSHWRPLEQIRL 29
          |||||:|:|:|
Db       137 fcrypswhrpldqlnl 152

RESULT 3
AAG83271
ID      AAG83271 standard; Protein; 691 AA.
XX
AC      AAG83271;
XX
DT      05-SEP-2001 (first entry)
XX
DE      Chlamydia trachomatis PmpA(C-term) fusion protein.
XX
KW      Chlamydia; vaccine; infection; fusion protein; antigen;
KW      pelvic inflammatory disease; trachoma; atherosclerosis; heart disease;
KW      acute respiratory tract infection; Cap1; CT529; OMCB;
KW      polymorphic membrane protein; pmp; thiol specific antioxidant; TSA.
XX
OS      Chlamydia trachomatis.
XX
PN      WO200140474-A2.
XX
PD      07-JUN-2001.
XX
PF      04-DEC-2000; 2000WO-US32919.
XX
PR      03-DEC-1999; 99US-0454684.
PR      19-APR-2000; 2000US-0556877.
PR      20-JUN-2000; 2000US-0598419.
XX
PA      (CORI-) CORIXA CORP.
XX
PI      Probst P, Bhatia A, Skelky YAW, Fling SP, Scholler J;
XX
DR      WPI; 2001-374831/39.
XX
CC      Chlamydia polypeptides and fusion proteins useful for preventing pelvic
XX      inflammatory disease, trachoma, acute respiratory tract infections,
XX      atherosclerosis and heart disease -
XX
PS      Claim 70; Page 267-268; 295pp; English.
XX
CC      The present sequence is provided in a specification relating to
CC      compounds and methods for the treatment and diagnosis of chlamydial
CC      infection. The compounds provided include polypeptides and fusion
CC      proteins comprising immunogenic portions of Chlamydia antigens
CC      and DNA sequences encoding such polypeptides. They are useful for
CC      vaccinating against chlamydial infection, which causes pelvic
CC      inflammatory disease, trachoma, acute respiratory tract infections,
CC      atherosclerosis and heart disease.
XX
SQ      Sequence 691 AA;

Query Match      30.7%; Score 80.5; DB 22; Length 691;
Best Local Similarity 50.0%; Pred. No. 0.053;
Matches 17; Conservative 1; Mismatches 15; Indels 1; Gaps 1;

QY      14 FCRYPSHWRPLEQIRLLTKPERRLSMLPLPSNN 47
          |||||:|:|:|
Db       137 fcrypswhrplnslnf-dpymalsassspqln 169

```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd

OM protein - protein search, using sw model

Run on: January 6, 2002, 09:33:18 ; Search time 135.97 Seconds
(without alignments)
50.561 Million cell updates/sec

Title:	US-09-897-042-16
Perfect score:	262
Sequence:	1 MNLVPMNPLVMNGFCRYPH.....RLTLKPERRLSWLPPLSNN 47

Scoring table: BLOSUM62

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

```

minimum DB seq length: 0
maximum DB seq length: 20000000000

```

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

```

1:  sp_archaea:17:*
2:  sp_bacteria:*
3:  sp_fungi:*
4:  sp_human:*
5:  sp_invertebrate:*
6:  sp_mammal:*
7:  sp_mhc:*
8:  sp_ornithelle:*
9:  sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	114.5	43.7	90	2	047336	047336 escherichia
2	83.5	31.9	217	2	053022	053022 escherichia
3	68	26.0	2081	1	09HU37	09HU37 thermoplasma
4	60.5	23.1	132	4	09NWS2	09NWS2 homo sapiens
5	59	22.5	132	4	09BYV8	09BYV8 homo sapiens
6	59	22.5	215	11	09CS38	09CS38 mus musculus
7	59	22.5	216	11	09DAY6	09DAY6 mus musculus
8	59	22.5	216	11	09D064	09D064 mus musculus
9	59	22.5	216	11	09D053	09D053 mus musculus
10	59	22.5	216	11	09D033	09D033 mus musculus
11	58	22.1	395	12	056762	056762 dengue virus
12	58	22.1	396	12	010247	010247 dengue virus
13	58	22.1	399	12	010248	010248 dengue virus
14	58	22.1	399	12	010254	010254 dengue virus
15	58	22.1	400	12	010255	010255 dengue virus
16	58	22.1	483	10	09FT55	09FT55 arabidopsis
17	57	21.8	420	4	014979	014979 homo sapiens
18	55.5	21.2	281	1	09V217	09V217 pyrococcus
19	55.5	21.2	311	9	09MCC8	09MCC8 lactobacillus

20	55	11.0	1327	5	09VLN6	09VLN6 drosophila
21	54.5	20.8	279	1	0580Z1	0580Z1 pyrococcus
22	54.5	20.8	362	2	08K9Y8	08K9Y8 bacillus ha
23	54.5	20.8	353	2	09UXM7	09UXM7 neisseria m
24	54.5	20.8	558	2	09JWB6	09JWB6 neisseria m
25	54.5	20.8	699	10	09AVC8	09AVC8 pistia sativ
26	54	20.6	123	2	052893	052893 coxiella bu
27	54	20.6	152	2	045954	045954 coxiella bu
28	54	20.6	325	2	025822	025822 helicobacte
29	54	20.6	378	11	09DAZ7	09DAZ7 mus musculu
30	54	20.6	387	12	010Z51	010Z51 dengue viru
31	54	20.6	581	11	035120	035120 mus musculu
32	54	20.6	581	11	09GYT7	09GYT7 mus musculu
33	54	20.6	612	8	0960Y3	0960Y3 diplophos t
34	53.5	20.4	185	8	033889	033889 adela trigr
35	53	20.2	332	2	051888	051888 prevotella
36	53	20.2	380	2	068873	068873 myxococcu
37	53	20.2	925	12	071101	071101 bovine aden
38	53	20.2	1215	1	09UWR7	09UWR7 pyrococcu
39	53	20.2	7829	5	018559	018559 caenorhabdi
40	52.5	20.0	227	8	09T7A8	09T7A8 anopheles m
41	52.5	20.0	452	2	053659	053659 streptomyce
42	52.5	20.0	485	11	068738	068738 mus musculu
43	52	19.8	411	4	060643	060643 homo sapien
44	52	19.8	463	10	09LJ48	09LJ48 arabidopsis
45	52	19.8	528	4	09NO62	09NO62 homo sapien
46	52	19.8	534	4	09NO61	09NO61 homo sapien
47	52	19.8	757	5	09VLI5	09VLI5 drosophila
48	52	19.8	805	5	09UJZ6	09UJZ6 drosophila
49	52	19.8	837	2	09HTE2	09HTE2 pseudomonas
50	52	19.8	921	11	09QX38	09QX38 rattus norv

ALIGNMENTS

RESULT	1			
ID	Q47336	PRELIMINARY;	PRT;	90 AA.
AC	Q47336			
DT	01-NOV-1996 (TREMBLrel. 01, Created)			
DT	01-NOV-1996 (TREMBLrel. 01, Last sequence update)			
DT	01-NOV-1999 (TREMBLrel. 12, Last annotation update)			
DE	LACZ-ALPHA PEPTIDE.			
OS	Escherichia coli.			
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;			
OC	Escherichia.			
OX	NCBI_Taxid=562;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Maneewanakul K., Maneewanakul S., Ippen-Ihler K.;			
RL	Submitted (AUG-1991) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; M74750; AAA24056.1; -			
DR	HSSP; P00722; 18GL.			
SO	SEQUENCE 90 AA; 10171 MW; 822AF7A109511DC CRC64;			
Query Match	43.7%;	Score 114.5;	DB 2;	Length 90;
Best Local Similarity	48.3%;	Pred. No. 8.8e-08;		
Matches 28;	Conservative 48;	Mismatches 5;	Indels 17;	Gaps 3
QY	7 NPLV--MNGFCRRP--SHWRPLEQ-----IRLTKPERRLSWLLPPLSNN 47			
Db	33 NPGVTQLNRLAHNPFAASWRNSEAEARTDRPSQQLRSLRLTKPERKLSWLLPPLSNN 90			
RESULT	2			
ID	053022	PRELIMINARY;	PRT;	217 AA.
AC	053022			
DT	01-JUN-1998 (TREMBLrel. 06, Created)			
DT	01-JUN-1998 (TREMBLrel. 06, Last sequence update)			
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)			

DE CLONING VECTOR PZERO-2T.
 GN FUSED-CCDB.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98172738; PubMed=9511747;
 RA Gabant P., Szpirer C.Y., Couturier M., Faelan M.;
 RT "Direct selection cloning vectors adapted to the genetic analysis of
 gram-negative bacteria and their plasmids."
 RL Gene 207:87-92(1998)
 DR EMBL; Y10545; CAA71575.1; -
 DR HSSP; P05703; 3VUB.
 DR InterPro; IPR002712; Codb.
 DR Pfam; PF01845; Codb; 1.
 DR Prodom; PD012578; Codb; 1.
 Q SEQUENCE 217 AA; 24893 MW; 8DCFF1B864228035 CRC64;

Query Match 31.9%; Score 83.5; DB 2; Length 217;
 Best Local Similarity 62.1%; Pred. No. 0.0027;
 Matches 18; Conservative 2; Mismatches 4; Indels 5; Gaps 2;

OY 2 NLVP-MNPVWNG---FCRYPHMRPLE 25
 DB 21 SLVPSDPLVTAASVLEFCRYPHMRPLE 49

RESULT 3
 ID 09HJ37 PRELIMINARY; PRT; 2081 AA.
 AC 09HJ37;
 DT 01-MAR-2001 (TREMBlrel. 16, Created)
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
 DE CONSERVED HYPOTHETICAL MEMBRANE PROTEIN.
 GN TAI136.
 OS Thermoplasma acidophilum.
 OC Archaea; Euryarchaeota; Thermoplasmatales; Thermoplasmataceae;
 CC Thermoplasma.
 OX NCBI_TaxID=2303;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=DSM 1728;
 RX MEDLINE=20479972; PubMed=11029001;
 RA Ruepp A., Graml W., Santos-Martinez M.-L., Koretke K.K., Volker C.,
 RA Mewes H.-W., Frishman D., Stocker S., Lupas A.N., Baumeister W.;
 RT "The genome sequence of the thermophilic scavenger Thermoplasma
 acidophilum."
 RL Nature 407:508-513(2000).
 DR EMBL; AL445066; CAC12262.1; -
 KW Complete proteome.
 SO SEQUENCE 2081 AA; 227496 MW; 6E02AA6470DB2EBD CRC64;

Query Match 26.0%; Score 68; DB 1; Length 2081;
 Best Local Similarity 38.6%; Pred. No. 3.1;
 Matches 17; Conservative 7; Mismatches 16; Indels 4; Gaps 2;

OY 4 VPMPPLVNGFCRYPHMR--PLEQIRLLTKPERRLSWLPLP 45
 DB 850 VPPNP--YNNVTAPSAWKIVPIDEAVTLTSEGGKQAQLPLPLS 891

RESULT 4
 ID 09NWS2 PRELIMINARY; PRT; 132 AA.
 AC 09NWS2;
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)

DE CDNA FLJ20640 FIS, CLONE KAT02911.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Watanabe K., Kumagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,
 RA Suzuki Y., Ohtsuka M., Nishi T., Shibahara T., Tanaka T.,
 RA Nakamura Y., Isono T., Sugano S.;
 RT "NDO human cDNA sequencing project."
 RL Submitted (FE8-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK006447; BA91305.1; -
 Q SEQUENCE 132 AA; 14013 MW; D9E8CB76D4920AF2 CRC64;

Query Match 23.1%; Score 60.5; DB 4; Length 132;
 Best Local Similarity 37.5%; Pred. No. 1.8;
 Matches 15; Conservative 5; Mismatches 19; Indels 1; Gaps 1;

OY 4 VPMPPLVNGFCRYPHMR--PLEQIRLLTKPERRLSWLPLP 42
 DB 12 VPSNPPLHRSFPCPRGMSHPVPTRELLERPAQADLLP 51

RESULT 5
 ID 09BV8 PRELIMINARY; PRT; 132 AA.
 AC 09BV8;
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE HYPOTHETICAL 14.0 KDA PROTEIN.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=CERVIX CARCINOMA;
 RA Strausberg R.;
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC000890; AAH00890.1; -
 KW Hypothetical protein.
 SO SEQUENCE 132 AA; 13997 MW; B5E8DF6395C61FB3 CRC64;

Query Match 22.5%; Score 59; DB 4; Length 132;
 Best Local Similarity 39.0%; Pred. No. 2.8;
 Matches 16; Conservative 5; Mismatches 18; Indels 2; Gaps 2;

OY 4 VPMPPLVNGFCRYPHMR--PLEQIRLLTKPERRLSWLPLP 43
 DB 12 VPSNPPLHRSFPCPRGMSHPVPTRELLERPAQADLLP 51

RESULT 6
 ID 09CS38 PRELIMINARY; PRT; 215 AA.
 AC 09CS38;
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE 1600012F09RIK PROTEIN (FRAGMENT).
 GN 1600012F09RIK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=EMBRYO;
 RX MEDLINE=21085660; PubMed=11217851;

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 6, 2002, 09:34:40 ; Search time 68.88 Seconds
(without alignments)
25.018 Million cell updates/sec

Title: US-09-897-042-16

Perfect score: 262
Sequence: 1 MNLVPMNPLVMNGFCRPSH.....RLTKPERRLSWLLPLPLSNN 47

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 segs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 50 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	55	21.0	421	1 CRT2_BOVIN	P42918 bos taurus
2	55	21.0	455	1 HMC2_BLAG	P54870 blatella g
3	54	20.6	970	1 T3RE_BPTI	P08764 bacterioph
4	53.5	20.4	123	1 VST1_HEVME	Q03499 hepatitis e
5	53.5	20.4	255	1 RCCEM_RHOPH	P51751 rhodospiril
6	53	20.2	309	1 AGAR_STPCO	P07883 streptomyc
7	53	20.2	911	1 HEX_ADEB3	P03278 bovine aden
8	52.5	20.0	282	1 DRN1_HUMAN	P24855 homo sapien
9	52	19.8	400	1 CRT1_BOVIN	P52153 bos taurus
10	52	19.8	416	1 CRTC_MOUSE	P14211 mus musculu
11	52	19.8	416	1 CRTC_RAT	P18418 rattus norv
12	52	19.8	417	1 CRTC_HUMAN	P27797 homo sapien
13	52	19.8	418	1 CRTC_RABIT	P51253 oryctolagus
14	52	19.8	922	1 NRPI_MOUSE	P09439 rattus norv
15	52	19.8	923	1 NRPI_RAT	P09373 mus musculu
16	52	19.8	1356	1 VCAP_VZVD	P09245 varicella-z
17	51.5	19.7	375	1 RND_ECOLI	P09155 escherichia
18	51.5	19.7	953	1 PODK_FLABI	Q39735 flavaria bi
19	51.5	19.7	953	1 PODK_FLATR	P22221 flavaria tr
20	51.5	19.7	1162	1 LEPR_MOUSE	P48336 mus musculu
21	51.5	19.7	1286	1 NGCA_CHICK	Q03696 gallus gall
22	51.5	19.7	4829	1 BIRE_HUMAN	Q09109 homo sapien
23	51	19.5	278	1 RHAS_ECOLI	P09377 escherichia
24	51	19.5	923	1 NRP1_HUMAN	O14786 homo sapien
25	50.5	19.3	184	1 YCDY_ECOLI	P75915 escherichia
26	50.5	19.3	277	1 RHAS_SALTY	P77029 salmonella
27	50.5	19.3	2182	1 CAB1_RAT	O88400 rattus norv
28	50	19.1	389	1 MDLA_BOCAL	P57551 buchnera ap
29	50	19.1	607	1 LPXB_CHLTR	O84416 chlamydia t
30	49.5	18.9	322	1 CRUB_TRICY	Q03443 tripedalia
31	49.5	18.9	359	1 CKR3_MOUSE	P51678 mus musculu
32	49.5	18.9	405	1 B3AR_CAPIH	O9xt57 capra hircu
33	49.5	18.9	469	1 TRME_SYREL	Q93910 synechococ

34	49.5	18.9	644	1 YEAG_ECOLI	P77391 escherichia
35	49.5	18.9	985	1 AGLU_ASPOR	Q12558 aspergillus
36	49	18.7	610	1 PTMA_BACSU	P42956 bacillus su
37	49	18.7	2054	1 YCF2_PINTH	P41653 pinus thunb
38	48.5	18.5	228	1 COX2_CUROU	P50693 culx quing
39	48.5	18.5	285	1 PLCA_MOUSE	Q35083 mus musculu
40	48.5	18.5	297	1 MT04_AQOAE	O66479 aquilex aeo
41	48.5	18.5	405	1 B3AR_BOVIN	P46626 bos taurus
42	48.5	18.5	446	1 AP50_SCHPO	O09731 schizosacch
43	48.5	18.5	493	1 XYLB_HARIN	P44401 haemophilus
44	48.5	18.5	587	1 TALA_BPDV	P13894 budgerigar
45	48.5	18.5	1026	1 EX5B_CHLTR	O84645 chlamydia t
46	48.5	18.5	1162	1 LEPR_RAT	O62959 rattus norv
47	48	18.3	331	1 PLIB_AGRBL	O93233 aglistrodon
48	48	18.3	468	1 IL9R_MOUSE	O01114 mus musculu
49	48	18.3	504	1 Y795_MENJA	O58205 methanococ
50	48	18.3	532	1 PUR2_ARATH	P52420 arabidopsis

ALIGNMENTS

```

RESULT 1
CRT2_BOVIN
ID CRT2_BOVIN STANDARD; PRT: 421 AA.
AC P42918;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE CALRETIICULIN, BRAIN ISOFORM 2 PRECURSOR (CRP5) (CALREGULIN) (HACBP).
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=93385184; PubMed=8373827;
RA Liu N., Fine R.E., Johnson R.J.;
RT "Comparison of cDNAs from bovine brain coding for two isoforms of
RT calreticulin."
RL Biochim. Biophys. Acta 1202:70-76(1993).
RE -1- FUNCTION: THIS PROTEIN BINDS CALCIUM. THERE ARE BOTH HIGH AND
RE LOW AFFINITY CALCIUM-BINDING SITES.
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: ENDOPLASMIC RETICULUM LUMEN.
CC -1- SIMILARITY: BELONGS TO THE CALRETIICULIN FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L13462; AAC37307.1;
DR InterPro: IPR001580; Calreticulin.
DR InterPro: IPR000866; ER_target.
DR Pfam: PF00262; calreticulin; 1.
DR PRINTS: PR00626; CALRETIICULIN.
DR PRODOM: PD001866; Calreticulin; 1.
DR PROSITE: PS00014; ER_TARGET; 1.
DR PROSITE: PS00803; CALRETIICULIN.
DR PROSITE: PS00804; CALRETIICULIN_2; 1.
DR PROSITE: PS00805; CALRETIICULIN_REPEAT; 3.
DR Endoplasmic reticulum; Calcium-binding; Repeat; signal.
KW SIGNAL
FT CHAIN 1 34
FT CHAIN 35 421
FT DOMAIN 35 201 N-DOMAIN.
FT DOMAIN 202 312 P-DOMAIN.
FT DOMAIN 313 421 C-DOMAIN.

```

Query Match 21.0%; Score 55; DB 1; Length 421;
Best Local Similarity 30.8%; Pred. No. 12;
Matches 12; Conservative 6; Mismatches 15; Indels 6; Gaps 2;

FT	DOMAIN	195	259	4 X APPROXIMATE REPEATS.
FT	REPEAT	195	206	1-1.
FT	REPEAT	214	225	1-2.
FT	REPEAT	231	242	1-3.
FT	REPEAT	248	259	1-4.
FT	DOMAIN	263	301	3 X APPROXIMATE REPEATS.
FT	REPEAT	263	273	2-1.
FT	REPEAT	277	287	2-2.
FT	REPEAT	291	301	2-3.
FT	DOMAIN	366	411	ASP/GLU/LYS-RICH.
FT	DISULFID	141	167	BY SIMILARITY.
FT	CARBOHYD	183	183	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	SITE	418	421	PREVENT SECRETION FROM ER.
SO	SEQUENCE	421 AA;	48812 MW;	0257E959F71528BC CRC64;

Query Match 21.0%; Score 55; DB 1; Length 421;
Best Local Similarity 30.8%; Pred. No. 12;
Matches 12; Conservative 6; Mismatches 15; Indels 6; Gaps 2;

FT <th>DOMAIN</th> <th>195</th> <th>259</th> <th>4 X APPROXIMATE REPEATS.</th>	DOMAIN	195	259	4 X APPROXIMATE REPEATS.
FT	REPEAT	195	206	1-1.
FT	REPEAT	214	225	1-2.
FT	REPEAT	231	242	1-3.
FT	REPEAT	248	259	1-4.
FT	DOMAIN	263	301	3 X APPROXIMATE REPEATS.
FT	REPEAT	263	273	2-1.
FT	REPEAT	277	287	2-2.
FT	REPEAT	291	301	2-3.
FT	DOMAIN	366	411	ASP/GLU/LYS-RICH.
FT	DISULFID	141	167	BY SIMILARITY.
FT	CARBOHYD	183	183	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	SITE	418	421	PREVENT SECRETION FROM ER.
SO	SEQUENCE	421 AA;	48812 MW;	0257E959F71528BC CRC64;

RESULT 2
HMC2_BLAGE
ID HMC2_BLAGE STANDARD; PRT; 455 AA.
AC PS4870:
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE HYDROXYMETHYLMUTARYL-COA SYNTHASE 2 (EC 4.1.3.5) (HMG-COA SYNTHASE 2)
GN HMGCS-2.
OS Blatella germanica (German cockroach).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Orthopteroidea; Dictyoptera; Blattaria;
OC Blattelloidea; Blattellidae; Blattellinae; Blattella.
ON NCBI_Taxid=6973;
RX MEDLINE=94216267; PubMed=7909314;
RA Blesa C., Martinez-Gonzalez J., Casals N., Haro D., Pluochs M.D.,
RA Biales X., Hegardt F.G.;
RT "Blattella germanica has two HMG-CoA synthase genes. Both are
regulated in the ovary during the gonadotrophic cycle."
RL J. Biol. Chem. 269:11707-11713(1994).
CC -1- FUNCTION: THIS ENZYME CONDENSES ACETYL-COA WITH ACETOACETYL-COA TO
FORM HMG-COA, WHICH IS THE SUBSTRATE FOR HMG-COA REDUCTASE (BY
SIMILARITY).
CC -1- CATALYTIC ACTIVITY: (S)-3-HYDROXY-3-METHYLMUTARYL-COA + COA =
ACETYL-COA + H(2)O + ACETOACETYL-COA.
CC -1- PATHWAY: PRODUCTION OF MEVALONATE FROM HMG-COA PRIOR TO THE
SYNTHESIS OF STEROLS AND ISOPRENOIDS SUCH AS JUVENILE HORMONES.
CC -1- SIMILARITY: BELONGS TO THE HMG-COA SYNTHASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X77516; CA56652.1; -;
DR InterPro: IPR000590; HMG_COA_synth.
DR Pfam: PF01154; HMG_COA_synth. 1.
DR PROSITE: PS01226; HMG_COA_SYNTHASE. 1.
KW Lyase; Sterol biosynthesis; Multigene family.
FT ACT_SITE 120 120
FT ACT_SITE 455 AA; 51425 MW; DD03035654863B1 CRC64;
SO SEQUENCE

Query Match 21.0%; Score 55; DB 1; Length 455;
Best Local Similarity 45.5%; Pred. No. 13;
Matches 10; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

Query Match 21.0%; Score 55; DB 1; Length 455;
Best Local Similarity 45.5%; Pred. No. 13;
Matches 10; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

FT	DOMAIN	195	259	4 X APPROXIMATE REPEATS.
FT	REPEAT	195	206	1-1.
FT	REPEAT	214	225	1-2.
FT	REPEAT	231	242	1-3.
FT	REPEAT	248	259	1-4.
FT	DOMAIN	263	301	3 X APPROXIMATE REPEATS.
FT	REPEAT	263	273	2-1.
FT	REPEAT	277	287	2-2.
FT	REPEAT	291	301	2-3.
FT	DOMAIN	366	411	ASP/GLU/LYS-RICH.
FT	DISULFID	141	167	BY SIMILARITY.
FT	CARBOHYD	183	183	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	SITE	418	421	PREVENT SECRETION FROM ER.
SO	SEQUENCE	421 AA;	48812 MW;	0257E959F71528BC CRC64;

Query Match 21.0%; Score 55; DB 1; Length 455;
Best Local Similarity 45.5%; Pred. No. 13;
Matches 10; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

FT	DOMAIN	195	259	4 X APPROXIMATE REPEATS.
FT	REPEAT	195	206	1-1.
FT	REPEAT	214	225	1-2.
FT	REPEAT	231	242	1-3.
FT	REPEAT	248	259	1-4.
FT	DOMAIN	263	301	3 X APPROXIMATE REPEATS.
FT	REPEAT	263	273	2-1.
FT	REPEAT	277	287	2-2.
FT	REPEAT	291	301	2-3.
FT	DOMAIN	366	411	ASP/GLU/LYS-RICH.
FT	DISULFID	141	167	BY SIMILARITY.
FT	CARBOHYD	183	183	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	SITE	418	421	PREVENT SECRETION FROM ER.
SO	SEQUENCE	421 AA;	48812 MW;	0257E959F71528BC CRC64;

Query Match 21.0%; Score 55; DB 1; Length 455;
Best Local Similarity 45.5%; Pred. No. 13;
Matches 10; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

FT	DOMAIN	195	259	4 X APPROXIMATE REPEATS.
FT	REPEAT	195	206	1-1.
FT	REPEAT	214	225	1-2.
FT	REPEAT	231	242	1-3.
FT	REPEAT	248	259	1-4.
FT	DOMAIN	263	301	3 X APPROXIMATE REPEATS.
FT	REPEAT	263	273	2-1.
FT	REPEAT	277	287	2-2.
FT	REPEAT	291	301	2-3.
FT	DOMAIN	366	411	ASP/GLU/LYS-RICH.
FT	DISULFID	141	167	BY SIMILARITY.
FT	CARBOHYD	183	183	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	SITE	418	421	PREVENT SECRETION FROM ER.
SO	SEQUENCE	421 AA;	48812 MW;	0257E959F71528BC CRC64;

C:Superfamily: calreticulin
C:Keywords: calcium binding; glycoprotein
F:1-34/Domain: signal sequence #status predicted <SIG>
F:35-421/Product: calreticulin, brain isoform 2 #status predicted <MAT>
F:418-421/Region: endoplasmic reticulum retention signal
F:141-167/Distulfide bonds: #status predicted
F:183/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 21.0%; Score 55; DB 2; Length 421;
Best Local Similarity 30.8%; Pred. No. 24;
Matches 12; Conservative 6; Mismatches 15; Indels 6; Gaps 2;

Qy 8 PLYNNGFCRYPSSHWRPLEQIRLLTKPERRLSWLPLPSN 46
Db 268 PLIQNP--EYKGEWKP-----RQIQNPKEKGIWHPEDIN 300

RESULT 3

hydroxymethylglutaryl-CoA synthase (EC 4.1.3.5) - German cockroach

C:Species: Blattella germanica (German cockroach)

C>Date: 07-Oct-1994 #sequence_revision 07-Oct-1994 #text_change 21-Jul-2000

C:Accession: A53565

R:Buesa, C.; Martinez-Gonzalez, J.; Casals, N.; Haro, D.; Pluach, M.D.; Belles, X.; He

J.; Biol. Chem. 269, 11707-11713, 1994

A:Title: Blattella germanica has two HMG-CoA synthase genes. Both are regulated in the c

A:Reference number: A53565; MUID:4216267

A:Accession: A53565

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-455 <BCE>

A:Cross-references: GB:W7516; NID:9488133; PIDN:CAA54652.1; PID:9488134

C:Superfamily: hydroxymethylglutaryl-CoA synthase

C:Keywords: carbon-carbon lyase; coenzyme A; oxo-acid-lyase

Query Match 21.0%; Score 55; DB 2; Length 455;
Best Local Similarity 45.5%; Pred. No. 26;
Matches 10; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

Qy 9 LVNMGFCRYPSSHWRPLEQIRLL 30
Db 269 LVNMFQYPERKYDQLQRL 290

RESULT 4

hypothetical protein PH0283 - Pyrococcus horikoshii

C:Species: Pyrococcus horikoshii

C>Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 21-Jul-2000

C:Accession: D71453

R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Halkawa, Y.; Hino, Y.; Yamamoto, S.; Seki

M.; Onfuku, Y.; Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.; Kishida, N.; Oguchi

DNA Res. 5, 55-76, 1998

A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic a

A:Reference number: A71000; MUID:98344137

A:Accession: D71453

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-279 <KAW>

A:Cross-references: GB:AP000001; NID:93236128; PIDN:BA29355.1; PID:93256672

A:Experimental source: strain OT3

A:Note: this accession replaces an interim accession for a sequence replaced by GenBank

C:Genetics:

A:Gene: PH0283

Query Match 20.8%; Score 54.5; DB 2; Length 279;
Best Local Similarity 27.9%; Pred. No. 17;
Matches 12; Conservative 10; Mismatches 14; Indels 7; Gaps 2;

Qy 4 VPNNPLV-----MNGFCRYPSSHWRPLEQIRLLTKPERRLSWL 40

Db 11 IPLNITRSEIHQESLLFATLFRP-EVIELIDPAPERLTMV 52

RESULT 5

cardamoyl-phosphate synthetase (glutamine subunit) pyrA [imported] - Bacillus halo

C:Species: Bacillus halodurans

C>Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 08-Dec-2000

C:Accession: A83967

R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; H

Nucleic Acids Res. 28, 4317-4331, 2000

A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a

A:Reference number: A83960; MUID:20263314

A:Accession: A83967

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-362 <STO>

A:Cross-references: GB:AP001515; GB:BA000004; NID:910174886; PIDN:BA06256.1; GSPDB:G

A:Experimental source: strain C-125

C:Genetics:

A:Gene: pyrA

C:Superfamily: cardamoyl-phosphate synthase (glutamine-hydrolyzing) small chain; carb

Query Match 20.8%; Score 54.5; DB 2; Length 362;
Best Local Similarity 34.1%; Pred. No. 23;
Matches 15; Conservative 7; Mismatches 13; Indels 9; Gaps 2;

Qy 6 MNP-----LVNMGFCRYPSSHWRPLEQIRLLTKPERRLSWLPLPS 45
Db 71 MNPAILGLIVEACDIPSMRSESLDLKAKO-----IPGLS 109

RESULT 6

hypothetical protein NMB1954 [imported] - Neisseria meningitidis (strain MC58 serogro

C:Species: Neisseria meningitidis

C>Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001

C:Accession: C81024

R:Nettelin, H.; Saunders, N.J.; Heldelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen,

Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.

ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masignani, V.; Pizze, M.

Science 287, 1809-1815, 2000

A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappelli, R.;

A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.

A:Reference number: A81000; MUID:20175755

A:Accession: C81024

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-553 <RPT>

A:Cross-references: GB:AE002543; GB:AE002098; NID:97227203; PIDN:AAE42283.1; PID:9722

A:Experimental source: serogroup B, strain MC58

C:Genetics:

A:Gene: NMB1954

Query Match 20.8%; Score 54.5; DB 2; Length 553;
Best Local Similarity 30.9%; Pred. No. 37;
Matches 17; Conservative 8; Mismatches 17; Indels 13; Gaps 3;

Qy 4 VPNNPLVNGFCR---YPSHWRPLEQIRLL-----TKPER---RLSWLPLPS 45
Db 281 LPALPLAVMYCKRILFSTGMLGVMALAVLVAVNFRQFDNLVWLLPLA 335

RESULT 7

probable inner membrane protein NMA0497 [imported] - Neisseria meningitidis (strain 2

C:Species: Neisseria meningitidis

C>Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001

C:Accession: F81967

R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Mo

GenCore version 4.5
Copyright (c) 1993 - 2000 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 6, 2002, 11:02:23 ; Search time 121.65 Seconds
(without alignments)
268,087 Million cell updates/sec

Title: US-09-897-042-15

Perfect score: 144

Sequence: 1 atgaactgtgacgcatgaa.....caccgtgagcaataactag 144

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_NA:*
1: /cgn2_6/prodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/prodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/prodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/prodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/prodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/prodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	108.2	75.1	3984	1	US-08-040-753-1
2	73.8	51.2	8501	3	US-08-793-900-1
3	71.2	49.4	2220	1	US-08-864-224-1
4	69.2	48.1	278	1	US-08-178-477B-33
5	68.6	47.6	6727	2	US-08-125-462-2
6	68.6	47.6	6727	2	US-08-891-848-2
7	68.4	47.5	4819	1	US-08-450-257-20
8	68.4	47.5	4819	1	US-08-450-246-20
9	68.4	47.5	4819	1	US-08-450-098-20
10	68.4	47.5	4819	1	US-08-451-233-20
11	68.4	47.5	4819	1	US-08-450-236-20
12	68.4	47.5	5574	1	US-08-450-257-22
13	68.4	47.5	5574	1	US-08-450-246-22
14	68.4	47.5	5574	1	US-08-450-098-22
15	68.4	47.5	5574	1	US-08-451-233-22
16	68.4	47.5	5574	1	US-08-450-236-22
17	68.2	47.4	375	3	US-08-870-370-7
18	68.2	47.4	375	3	US-08-870-370-8
19	68.2	47.4	375	3	US-08-870-370-9
20	68	47.2	1010	4	US-09-070-408-1
21	68	47.2	2735	2	US-08-920-095-1
22	68	47.2	2735	5	PCT-US96-05800-1
23	68	47.2	4411	2	US-08-929-967-5
24	68	47.2	5248	5	PCT-US96-05611A-21
25	68	47.2	5443	2	US-08-929-967-1
26	68	47.2	5616	2	US-08-929-967-3
27	68	47.2	6363	2	US-08-929-967-6

C	28	68	47.2	6619	2	US-08-683-007A-3	Sequence 3, Appl
	29	67.8	47.1	6873	4	US-09-131-028A-1	Sequence 1, Appl
C	30	67.8	47.1	6873	4	US-09-131-028A-8	Sequence 8, Appl
	31	67.8	47.1	8430	4	US-09-131-028A-6	Sequence 6, Appl
C	32	67.8	47.1	8430	4	US-09-131-028A-10	Sequence 10, Appl
	33	67.4	46.8	7659	4	US-09-128-314-4	Sequence 4, Appl
	34	67.4	46.8	8157	4	US-09-128-314-3	Sequence 3, Appl
	35	67	46.5	111	2	US-08-606-288-9	Sequence 9, Appl
	36	67	46.5	111	3	US-09-347-483-9	Sequence 9, Appl
	37	67	46.5	234	2	US-08-687-865A-20	Sequence 9, Appl
	38	67	46.5	234	4	US-09-043-711-20	Sequence 20, Appl
	39	67	46.5	600	2	US-08-125-462-1	Sequence 20, Appl
	40	67	46.5	600	2	US-08-891-848-1	Sequence 1, Appl
	41	67	46.5	2029	4	US-09-136-574A-46	Sequence 1, Appl
	42	67	46.5	4410	1	US-08-594-469-1	Sequence 46, Appl
	43	67	46.5	4410	2	US-08-906-957-1	Sequence 1, Appl
	44	67	46.5	4977	1	US-08-450-257-14	Sequence 14, Appl
	45	67	46.5	4977	1	US-08-450-257-17	Sequence 17, Appl

ALIGNMENTS

RESULT 1
US-08-040-753-1
: Sequence 1, Application US/08040753
: Patent No. 5464745
: GENERAL INFORMATION:
: APPLICANT: Mierendorf, Robert
: APPLICANT: Garber, Richard
: APPLICANT: No. 5464745y, Robert
: APPLICANT: Hammer, Beth
: TITLE OF INVENTION: Protein Ligand Binding
: TITLE OF INVENTION: Region Mapping System
: NUMBER OF SEQUENCES: 2
: CORRESPONDENCE ADDRESSES:
: ADDRESSEE: Quarles and Brady
: STREET: 1 South Pinckney St., Box 2113
: CITY: Madison
: STATE: WI
: COUNTRY: USA
: ZIP: 53701-2113
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/040.753
: FILING DATE: 19930331
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Seay, Nicholas J
: REGISTRATION NUMBER: 27386
: REFERENCE/DOCKET NUMBER: 70-399-9001-1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 608-251-5000
: TELEFAX: 608-251-9166
: TELEX:
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3984 base pairs
: TYPE: NUCLEIC ACID
: STRANDEDNESS: double
: TOPOLOGY: circular
: MOLECULE TYPE: DNA (genomic)
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: IMMEDIATE SOURCE:
: CLONE: pTOPE-1b(+)
: FEATURE:
: NAME/KEY: promoter
: LOCATION: 67..85

GenCore version 4.5
Copyright (c) 1993 - 2000 Compen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 6, 2002, 11:06:50 ; Search time 256.95 Seconds

(without alignments)
480.462 Million cell updates/sec

Title: US-09-897-042-15

Perfect score: 144
Sequence: 1 atgactgtgacccatgaa.....caccgtgacataactag 144

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

1: N.Geneseq_1101.*
2: /SID2/gcgdata/geneseq/geneseq/NA1980.DAT.*
3: /SID2/gcgdata/geneseq/geneseq/NA1981.DAT.*
4: /SID2/gcgdata/geneseq/geneseq/NA1982.DAT.*
5: /SID2/gcgdata/geneseq/geneseq/NA1983.DAT.*
6: /SID2/gcgdata/geneseq/geneseq/NA1984.DAT.*
7: /SID2/gcgdata/geneseq/geneseq/NA1985.DAT.*
8: /SID2/gcgdata/geneseq/geneseq/NA1986.DAT.*
9: /SID2/gcgdata/geneseq/geneseq/NA1987.DAT.*
10: /SID2/gcgdata/geneseq/geneseq/NA1988.DAT.*
11: /SID2/gcgdata/geneseq/geneseq/NA1989.DAT.*
12: /SID2/gcgdata/geneseq/geneseq/NA1990.DAT.*
13: /SID2/gcgdata/geneseq/geneseq/NA1991.DAT.*
14: /SID2/gcgdata/geneseq/geneseq/NA1992.DAT.*
15: /SID2/gcgdata/geneseq/geneseq/NA1993.DAT.*
16: /SID2/gcgdata/geneseq/geneseq/NA1994.DAT.*
17: /SID2/gcgdata/geneseq/geneseq/NA1995.DAT.*
18: /SID2/gcgdata/geneseq/geneseq/NA1996.DAT.*
19: /SID2/gcgdata/geneseq/geneseq/NA1997.DAT.*
20: /SID2/gcgdata/geneseq/geneseq/NA1998.DAT.*
21: /SID2/gcgdata/geneseq/geneseq/NA1999.DAT.*
22: /SID2/gcgdata/geneseq/geneseq/NA2000.DAT.*
23: /SID2/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	108.2	75.1	3984	16	Vector plasmid pTO
2	83.2	57.8	1104	20	Mycobacterium tube
3	80.4	55.8	8031	21	Human Lung cancer
4	79.4	55.1	5502	22	Human Lung tumour
5	79.2	55.0	8970	21	Nucleotide sequenc
6	79	54.9	3572	20	M. tuberculosis an
7	79	54.9	3572	20	M. tuberculosis re
8	79	54.9	7507	21	M. tuberculosis ge
9	79	54.9	7676	19	M. tuberculosis fu
10	79	54.9	7676	19	Mycobacterium anti
11	79	54.9	7676	20	Mycobacterium tube

12	79	54.9	7676	20	AA19368	M. tuberculosis fu
13	79	54.9	7676	20	AA19156	M. tuberculosis fu
14	79	54.9	8217	22	AA503795	M. tuberculosis fu
15	75	52.1	945	15	AA05302	gp55-Asn-Gly-Pro-h
16	73.8	51.2	8501	17	AA12907	Novel expression p
17	73.8	51.2	8793	22	AA503794	M. tuberculosis ve
18	71.8	49.9	5770	21	AA15043	Nucleotide sequenc
19	71.4	49.6	5870	21	AA15044	Nucleotide sequenc
20	71.2	49.4	5220	21	AA58064	Vector plasmid pUN
21	70.4	48.9	5216	20	AA29548	PK90 TIR coupling
22	70	48.6	1237	21	AA53600	M. tuberculosis an
23	69.8	48.5	363	19	AAV44696	pTRP cassette in
24	69.8	48.5	5142	22	AAE29645	Plasmid pLTMF1Y1
25	69.8	48.5	5230	20	AA07549	PTREX1 mutant beta
26	69.8	48.5	5230	21	AA15041	Nucleotide sequenc
27	69.8	48.5	5497	22	AAE29646	Plasmid pTRMFL
28	69.8	48.5	5906	21	AA15042	Nucleotide sequenc
29	68.6	47.6	6727	20	AA09858	pET-11d-ANG-E6 DNA
30	68.6	47.6	6727	20	AAE63616	Nucleic acid sequ
31	68.4	47.5	116	22	AAE82146	Plasmid pET-3a nuc
32	68.4	47.5	4819	15	AA04360	Plasmid pPR-delta
33	68.4	47.5	4819	18	AAE89768	Plasmid pPR-delta
34	68.4	47.5	4819	18	AAE88234	Plasmid pPR-delta
35	68.4	47.5	4819	18	AAE84576	Plasmid pPR-delta
36	68.4	47.5	4819	19	AAV6594	Plasmid pPR-delta
37	68.4	47.5	5231	20	AA07550	Plasmid pPR-delta
38	68.4	47.5	5574	15	AA04361	Plasmid pPR-delta
39	68.4	47.5	5574	18	AAE89770	Plasmid pPR-delta
40	68.4	47.5	5574	18	AAE88235	Plasmid pPR-delta
41	68.4	47.5	5574	18	AAE84578	Plasmid pPR-delta
42	68.4	47.5	5574	19	AAV6596	Plasmid pPR-delta
43	68.4	47.5	6499	21	AAV64015	Nucleotide sequenc
44	68.2	47.4	375	19	AAV54056	DNA fragment 2 of
45	68.2	47.4	375	19	AAV54057	DNA fragment 3 of

ALIGNMENTS

RESULT 1	AA07310	standard; DNA; 3984 BP.
ID	AA07310	
AC	AA07310	
XX		
DT	18-MAR-1996	(first entry)
XX		
DE	Vector plasmid pTOPE-1b(+).	
XX		
KW	Plasmid pTOPE-1b(+); vector; ligand binding domain; epitope mapping;	
KM	antigen; Escherichia coli; ds; cyclic.	
XX		
OS	Synthetic.	
XX		
FH	Key	Location/Qualifiers
FT	promoter	67..85
FT		/*tag- a
FT		/*function- T7 promoter
FT		134..139
FT	RBS	/*tag- b
FT		148..1092
FT	CDS	/*tag- c
FT		/*product- 5' end of T7 gene 10 fusion protein
FT		cloning by insertion of putative epitope
FT		encoding DNA"
XX	US5464745-A.	
XX	07-NOV-1995.	
XX	31-MAR-1993;	9305-0040753.

XX	31-MAR-1993: 93US-0040753.
XX	(NOVA-) NOVAGEN INC.
PA	
XX	Garber R, Hammer B, Miendorf R, Novy R;
PI	
XX	WPI: 1995-392610/50.
DR	
XX	P-PsDB: AAR87027.
DR	
XX	
PT	Mapping ligand binding domains, esp. epitope(s), of proteins - by
PT	expressing peptide(s) encoded by random gene fragments and testing
PT	for ligand binding
XX	
PS	Disclosure: Column 11-16; 12pp: English.
XX	
CC	Vector plasmid pTOPE-1b(+) (AA07310) is a high-copy-number version
CC	of pBR322-derivative plasmid pET-17xb. It is used in a method for
CC	mapping the ligand binding domains (LBD), esp. antigenic regions, of
CC	a protein. Putative LBD-encoding DNA is inserted into the vector
CC	such that the LBD is expressed as a fusion to the T7 gene 10 leader
CC	(AAR87027). The fusion protein accumulates as inclusion bodies in
CC	<i>Escherichia coli</i> host cells and can be screened for its ability to
CC	bind a ligand.
XX	
SO	Sequence 3964 BP; 960 A; 1004 C; 1007 G; 1013 T; 0 other:

Query Match	75.1%	Score 108.2	DB 16	Length 3984
Best Local Similarity	97.3%	Pred. No. 7.7e-27		
Matches 110	Conservative	0	Mismatches 37	Indels 0
Oy	32	tgaagaaattctgcagatcatcatcaactgagcgccgctcgagcagatccgcgtctaa	91	
Db	980	tgcctgaattctgcagatcatcatcaactgagcgccgctcgagcagatccgcgtctaa	1039	
Oy	92	caaaagcccgaaagaaagctgaattgctgcgtctccacgcctgagcaataactg	144	
Db	1040	caaaagcccgaaagaaagctgaattgctgcgtctccacgcctgagcaataactg	1092	

RESULT	2
AAZ20204	
ID	AAZ20204.. standard; DNA; 1104 BP.
AC	
XX	
XX	AAZ20204:
DT	
DT	17-JAN-2000 (first entry)
XX	
DE	Mycobacterium tuberculosis antigen fusion protein Mtub36f DNA.
XX	
KW	Tuberculosis; antigen; fusion protein; Mtub36f; ERD14; DPV; MTI;
KW	diagnosis; therapy; vaccine; immunogen; ss.
XX	
OS	Mycobacterium tuberculosis.
XX	
PN	W09951748-A2.
PD	
PD	14-OCT-1999.
XX	
PF	07-APR-1999; 99WO-US07717.
XX	
PR	07-APR-1998; 98US-0056556.
PR	30-DEC-1998; 98US-0223040.
XX	
PA	(CORI-) CORIXA CORP.
XX	
PI	Skelky YAM, Alderson M, Campos-Neto A;
XX	
DR	WPI: 1999-601610/51.
DR	P-RSDB: AA132069.
PT	
PT	

XX Example: Fig 11A-B; 83bp; English.
PS
XX
CC This DNA sequence includes a coding region for a recombinant
CC Mycobacterium tuberculosis tri-antigen fusion protein (see AY32069)
CC termed M30f, composed of the antigens ERD1, DPV and M71. The
CC DNA is useful for the recombinant production of the fusion protein.
CC Coding sequences for the antigens were modified by PCR in order
CC to facilitate their fusion and subsequent expression of the fusion
CC protein, and were then ligated to obtain the present polynucleotide.
CC The invention provides fusion proteins (see AY32057-71) containing
CC at least 2 M. tuberculosis antigens. The new fusion proteins and
CC polynucleotides encoding them are useful as vaccines for preventing
CC tuberculosis (claimed), for diagnosis (via in vitro assays or
CC intradermal skin tests for detection of anti-M. tuberculosis
CC antibodies), monitoring of disease progression, and treatment of
CC tuberculosis. They are more effective immunogens than mixtures of
CC the individual protein components.
XQ
XQ Sequence 1104 BP; 216 A; 347 C; 333 G; 208 T; 0 other;

Query Match	Similarity	Score	DB	Length
Best Local	85.0%	96.6%	Pred. No. 1.4e-18;	1104
Matches	85;	Conservative	0;	Mismatches 3; Indels 0; Gaps 0;

QY	32	tgaagaaatttcgcagatattcattcaactgtgcgcgcgtctgcagatccgcgtctaa	91
Db	1017	tgtcgaatttcgcagatattcattcaactgtgcgcgcgtctgcagatccgcgtctaa	1076

OY	92	caaaagcccgaaaggaagctgaagttggct	119
Db	1077	caaaagcccgaaaggaagctgaagttggct	1104

xx	RESULT	3
xx	AAC65950	
xx	ID	AAC65950 standard; DNA; 8031 BP.
xx	AC	AAC65950;
xx	DT	21-FEB-2001 (first entry)
xx	DE	Human lung cancer-associated DNA for L762P expression construct.
xx	XX	
xx	KW	Lung cancer; therapy; treatment; human; tumor; immunogenic; cytostatic;
xx	RN	vaccine; detection; ss.
xx	OS	Homo sapiens.
xx	PN	WO200061612-A2.
xx	PD	19-OCT-2000.
xx	PF	03-APR-2000; 2000WO-US08896.
xx	PR	02-APR-1999; 99US-0285479.
xx	PR	17-DEC-1999; 99US-0466396.
xx	PR	30-DEC-1999; 99US-0476496.
xx	PR	10-JAN-2000; 2000US-0480884.
xx	PR	22-FEB-2000; 2000US-0510376.
xx	PA	(CORI-) CORIXA CORP.
xx	PI	Wang T, Fan L;
xx	DR	WPI; 2000-628399/60.
xx	PT	Isolated polypeptide comprising an immunogenic portion of a lung tumor
xx	PT	protein is used for detecting and monitoring progression of lung cancer
xx	PT	in a patient -
xx	PS	Claim 1a; Page 211-213; 261pp; English.

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 6, 2002, 10:33:27 ; Search time 2642.04 Seconds
(without alignments)
585.681 Million cell updates/sec

Title: US-09-897-042-15

Perfect score: 144

Sequence: 1 atgaactgtgaccatgaa.....caccgtgagcaataactag 144

Scoring table:

IDENTITY NUC
Gapop 10.0 / Gapext 1.0

Searched: 11351937 segs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_hic:*
10: qb_est1:*
11: qb_est2:*
12: qb_hic:*
13: qb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rtd:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	56	38.9	602	13	B06973
2	51.6	35.8	301	11	BF286760
3	51.6	35.8	301	11	BF286761
4	49	34.0	594	10	BE439485
5	48.2	33.5	333	10	AU055647
6	48	33.3	574	10	BE439542
7	48	33.3	748	10	BE439434
8	47.2	32.8	245	11	BF286756
9	45.4	31.5	359	11	BF286758
10	44.8	31.1	302	11	BF286757
11	44.8	31.1	348	10	BE035760
12	44.4	30.8	220	11	BF286794

C 13	44.4	30.8	327	11	BF286781	BF286781	EST451372
C 14	43.6	30.3	332	11	BF286778	BF286778	EST451369
C 15	43.6	30.3	332	11	BF286780	BF286780	EST451371
C 16	43.6	30.3	713	13	A2574086	A2574086	324PEF10
C 17	43.2	30.0	722	13	A2570378	A2570378	273PEF12
C 18	43.2	30.0	761	13	A2572147	A2572147	301PEG11
C 19	43	29.9	247	11	BF286712	BF286712	EST451303
C 20	43	29.9	247	11	BF286714	BF286714	EST451305
C 21	42.8	29.7	184	11	BF286797	BF286797	EST451388
C 22	42.8	29.7	378	11	BF286772	BF286772	EST451363
C 23	42.6	29.6	234	10	A1110514	A1110514	394 neona
C 24	42.6	29.6	734	10	AM231897	AM231897	2A30N6 Su
C 25	42.6	29.6	891	10	BE205538	BE205538	AOB64 On1
C 26	42.4	29.4	272	10	AU055688	AU055688	AU055688
C 27	42.4	29.4	287	11	BF286768	BF286768	EST451359
C 28	42.4	29.4	430	11	BF286771	BF286771	EST451362
C 29	42.4	29.4	479	10	AA545726	AA545726	HBMSP2E4-
C 30	42.4	29.4	861	13	AF307808	AF307808	AF307808
C 31	42.2	29.3	141	11	BF286696	BF286696	EST451287
C 32	42.2	29.3	141	11	BF286749	BF286749	EST451340
C 33	42.2	29.3	300	11	BF286774	BF286774	EST451365
C 34	42.2	29.3	346	10	BE035761	BE035761	MO16E08 M
C 35	42.2	29.3	356	10	BE035762	BE035762	MO16E09 M
C 36	42.2	29.3	361	10	AM697836	AM697836	EST0019 C
C 37	42.2	29.3	565	11	BI273540	BI273540	F19A1 Fee
C 38	42.2	29.3	570	11	BG577408	BG577408	N120 S5H7
C 39	42.2	29.3	624	10	AJ410657	AJ410657	AJ410657
C 40	42.2	29.2	130	11	BG673766	BG673766	906 C11a
C 41	42.2	29.2	183	10	AU055645	AU055645	AU055645
C 42	42.2	29.2	206	10	AA389640	AA389640	HPBEST- H
C 43	42.2	29.2	221	10	AA389641	AA389641	HBCEST- H
C 44	42.2	29.2	247	10	AM608990	AM608990	PM3-PT004
C 45	42.2	29.2	257	13	AF034178	AF034178	AF034178

ALIGNMENTS

RESULT 1
B06973 602 bp DNA GSS 13-JUL-1996
LOCUS CSRL-8b6-t CSRL flow sorted Chromosome 11 specific cosmid Homo
DEFINITION sapiens genomic clone CSRL-8b6, DNA sequence.
ACCESSION B06973
VERSION B06973.1 GI:1416251
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Evans,G.A., Bubeck,D., Davies,C., Hahner,L., Oliver,T., Gilbert,M.,
Mammalia: Eutheria; Primates: Catarrhini; Homnidae; Homo.
1 (Bases 1 to 602)
AUTHORS Jones,D., Ward,T., Gillilan,E., Schagemann,J., Probst,S., Harris,
J., DeFord,J., McFarland,J., Burzinski,K., Khan,M., Kupfer,K. and
Garner,H.R.
Genomic Sequence Sampled Map of Chromosome 11
Unpublished (1996)
CONTACT: Evans GA, Shane Probst
McDermott Center for Human Growth and Development
University of Texas Southwestern Medical Center At Dallas
5323 Harry Hines Blvd, Dallas TX 75235-8591
Tel: 214-648-1600
Fax: 214-648-1666
Email: gjevans@utsw.swmed.edu, shane@mcdermott.swmed.edu
PCR PRIMERS
FORWARD: GAAGAAGAGACTATATCCG
BACKWARD: ATTGGTTAAAAAATGACGCG
Seq primer: T3
Class: cosmid ends
High quality sequence stop: 602.
Location/Qualifiers
1..602
/organism="Homo sapiens"

FEATURES
Source

/db_xref="taxon:9606"
/clone="CSRL-8b6"
/clone_lib="CSRL flow sorted Chromosome 11 specific cosmid"
/sex="female"
/cell_type="chimeric hamster somatic cell hybrid"
/note="Vector: SCOS-1; Human Chromosome 11 specific cosmid library prepared from flow sorted human Chromosome 11 derived from Chinese Hamster Ovary (CHO) monochromosomal somatic cell hybrid, J1"
BASE COUNT 148 a 126 c 151 g 143 t 34 others
ORIGIN

Query Match 38.9% Score 56; DB 13; Length 602;
Best Local Similarity 98.5% Pred. No. 1.7e-06;
Matches 67; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
Db 101 GATCGGCTGCTNACAAAGCCGAAAGAGCTGAGTGGCTGCTCCACGCTGAGCA 160
Gy 137 ataaatag 144
Db 161 ATAACTAG 168

RESULT 2
BF286760/c EST 28-NOV-2000
LOCUS BF286760 Rat Gene Index, normalized rat, Rattus norvegicus cDNA
DEFINITION BF286760 Rattus norvegicus cDNA clone RGIF57 3' sequence, mRNA sequence.
VERSION BF286760.1 GI:11217830
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE 1 (bases 1 to 301)
AUTHORS Malek,R.L., Cho,J., Lee,Y., Karamycheva,S., Parvizli,B., Pertea,G.,
Sultana,R., Tsai,J., White,J., Quackenbush,J. and Lee,N.H.
TITLE Generation of ESTs from Normalized Rat Embryo, Bento Soares
JOURNAL Unpublished (2000)
COMMENT Contact: Lee, NH
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org

This clone is available through the ATCC, contact the ATCC
tel#703-365-2700 for further information.
FEATURES
Location/Qualifiers
1..301

/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="RGIF57"
/clone_lib="Rat Gene Index, normalized rat, Rattus
norvegicus cDNA"
/tissue_type="mixed tissue"
/lab_host="DH5-alpha"
/note="Vector: pT37Pac; Site_1: EcoRI; Site_2: NotI;
Combination of ROV, RBR, RRI, RLI, RPL, RLU, RRM, RMU, RSP
RHE, RPC, RPN"
BASE COUNT 72 a 60 c 72 g 97 t
ORIGIN

Query Match 35.8% Score 51.6; DB 11; Length 301;
Best Local Similarity 67.9% Pred. No. 3e-05;
Matches 72; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

Gy 32 tgaacgaattctgcagatataccatcacctgagcgctgcagagatctcgctgctaa 91
Db 213 TGCCTGGAATTCCTGCAGATATCCATCAGCTGCGCGCCCTGAGCATGTCCAAATATTGA 154
Gy 92 caaagcccgaaagagagctgagctgctgcctccacgcctgagcaa 137
Db 153 AAAAACCACATGACAGCAAGACAGCTGCTCACTACATCATGAGGAA 108

RESULT 3
BF286761/c EST 28-NOV-2000
LOCUS BF286761 Rat Gene Index, normalized rat, Rattus norvegicus cDNA
DEFINITION BF286761 Rattus norvegicus cDNA clone RGIF58 3' sequence, mRNA sequence.
VERSION BF286761.1 GI:11217831
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE 1 (bases 1 to 301)
AUTHORS Malek,R.L., Cho,J., Lee,Y., Karamycheva,S., Parvizli,B., Pertea,G.,
Sultana,R., Tsai,J., White,J., Quackenbush,J. and Lee,N.H.
TITLE Generation of ESTs from Normalized Rat Embryo, Bento Soares
JOURNAL Unpublished (2000)
COMMENT Contact: Lee, NH
Other ESTs: EST451353
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org

This clone is available through the ATCC, contact the ATCC
tel#703-365-2700 for further information.
FEATURES
Location/Qualifiers
1..301

/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="RGIF58"
/clone_lib="Rat Gene Index, normalized rat, Rattus
norvegicus cDNA"
/tissue_type="mixed tissue"
/lab_host="DH5-alpha"
/note="Vector: pT37Pac; Site_1: EcoRI; Site_2: NotI;
Combination of ROV, RBR, RRI, RLI, RPL, RLU, RRM, RMU, RSP
RHE, RPC, RPN"
BASE COUNT 72 a 60 c 72 g 97 t
ORIGIN

Query Match 35.8% Score 51.6; DB 11; Length 301;
Best Local Similarity 67.9% Pred. No. 3e-05;
Matches 72; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

Gy 32 tgaacgaattctgcagatataccatcacctgagcgctgcagagatctcgctgctaa 91
Db 213 TGCCTGGAATTCCTGCAGATATCCATCAGCTGCGCGCCCTGAGCATGTCCAAATATTGA 154
Gy 92 caaagcccgaaagagagctgagctgctgcctccacgcctgagcaa 137
Db 153 AAAAACCACATGACAGCAAGACAGCTGCTCACTACATCATGAGGAA 108

RESULT 4
BE439485/c EST 25-JUL-2000
LOCUS BE439485 594 bp mRNA
DEFINITION HTML-116F1 HTML Homo sapiens cDNA, mRNA sequence.
ACCESSION BE439485
VERSION BE439485.1 GI:9438967
KEYWORDS EST.
SOURCE human.

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 6, 2002, 11:00:34 ; Search time 1586.23 Seconds
(without alignments)
1497.635 Million cell updates/sec

Title: US-09-897-042-15

Perfect score: 144
Sequence: 1 atgaacttgtagccgatgaa.....cacgcgtgagcaataactag 144

Scoring table: IDENTITY_NUC
Gapox 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl: *
1: gb_da: *
2: gb_hcg: *
3: gb_in: *
4: gb_om: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_sts: *
12: gb_sy: *
13: gb_un: *
14: gb_vi: *
15: em_ba: *
16: em_fun: *
17: em_hum: *
18: em_in: *
19: em_om: *
20: em_or: *
21: em_ov: *
22: em_pat: *
23: em_ph: *
24: em_pl: *
25: em_ro: *
26: em_sts: *
27: em_sy: *
28: em_un: *
29: em_vi: *
30: em_htgo_hum: *
31: em_htgo_inv: *
32: em_htgo_rod: *
33: em_htg_hum: *
34: em_htg_inv: *
35: em_htg_rod: *
36: em_htg_other: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	108.2	75.1	3984	6	I15353
2	79.4	55.02	5502	6	AX063158
3	75	52.1	717	6	A36847
4	75	52.1	945	6	A36849
5	73.8	51.2	8501	6	A49428
6	71.2	49.4	2039	12	AF156991
7	71.2	49.4	2060	12	AF156992
8	71.2	49.4	2075	12	AF156990
9	71.2	49.4	2220	6	AR067785
10	71.2	49.4	2220	12	AF143506
11	71.2	49.4	2326	12	AF143508
12	71.2	49.4	2490	12	AF143507
13	71.2	49.4	2526	12	AF149259
14	71.2	49.4	2547	12	AF149259
15	71.2	49.4	2562	12	AF149258
16	71.2	49.4	2669	12	AF149261
17	71.2	49.4	2788	12	AF149263
18	71.2	49.4	2811	12	AF149262
19	71.2	49.4	2966	12	AF149266
20	71.2	49.4	3028	12	AF149267
21	71.2	49.4	3378	12	AF149265
22	71.2	49.4	3417	12	AF149264
23	70.4	48.9	5217	6	AX001271
24	70.2	48.8	3123	12	PMB027D
25	70	48.6	3584	1	ECOLACAB
26	70	48.6	3741	12	PKSM711
27	70	48.6	5025	12	ECOT659
28	70	48.6	5041	12	PKSM713
29	70	48.6	5041	12	PKSM715
30	70	48.6	5071	12	PKSM710
31	70	48.6	5406	12	ECOT660
32	69.8	48.5	383	6	A61284
33	69.8	48.5	5142	6	AX069289
34	69.8	48.5	5230	6	AX001272
35	69.8	48.5	5505	6	AX069290
36	69.4	48.2	2863	12	AF013597
37	69.2	48.1	278	6	AR009724
38	69	47.9	2828	12	CVP09724
39	69	47.9	2829	12	CVP09724
40	69	47.9	2830	12	CVP09724
41	69	47.9	2831	12	CVP09724
42	68.6	47.6	2730	1	PET81P1P
43	68.6	47.6	2731	1	PET81P1P
44	68.6	47.6	2891	12	CVT7MCO
45	68.6	47.6	6101	12	CVP09724

ALIGNMENTS

RESULT 1
LOCUS I15353 3984 bp DNA
DEFINITION Sequence 1 from patent US 5464745.
ACCESSION I15353
VERSION I15353.1 GI:1250261
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 3984)
AUTHORS Mierendorf, R., Garber, R., Novy, R. and Hammer, B.
TITLE Protein ligand binding region mapping system
JOURNAL Patent: US 5464745-A 1 07-NOV-1995;
FEATURES
source Location/Qualifiers
1..3984
/organism="unknown"
BASE COUNT 960 a 1004 c 1007 g 1013 t
ORIGIN

PAT 02-APR-1996

Query Match 75.1%; Score 108.2; DB 6; Length 3984;
Best Local Similarity 97.3%; Pred. No. 9.7e-24;
Matches 110; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 32 tgaagaattctgcagatcatcactgctgcgcgtcagcagatcgcgtctaa 91
|||
DB 980 tgcctggaattctgcagatcatcactgctgcgcgtcagcagatcgcgtctaa 1039
|||

QY 92 caaagcccgaaagaaagctgagttgctctgcaccgctgagcaataactag 144
|||||
DB 1040 caaagcccgaaagaaagctgagttgctctgcaccgctgagcaataactag 1092
|||||

RESULT 2
AX063158 5502 bp DNA PAT 24-JAN-2001
LOCUS Sequence 785 from Patent WO0100828.
DEFINITION AX063158
ACCESSION AX063158
VERSION AX063158.1 GI:12541015
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE Wang, T., Bangur, C.S., Lodes, M.J., Fanger, G.R., Veddy, T.S.,
AUTHORS Carter, D., Retter, M.W. and Mannion, J.
TITLE Compositions and methods for the therapy and diagnosis of lung
cancer
JOURNAL Patent: WO 0100828-A 785 04-JAN-2001;
CORIXA CORPORATION (US)
FEATURES
source 1..5502
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT 1306 a 1479 c 1409 g 1308 t
ORIGIN

Query Match 55.1%; Score 79.4; DB 6; Length 5502;
Best Local Similarity 84.5%; Pred. No. 1.3e-14;
Matches 109; Conservative 0; Mismatches 1; Indels 19; Gaps 1;

QY 35 acgaattctgcagatcatcactgctgcgcgtcagc----- 76
|
DB 5305 atgaattctgcagatcatcactgctgcgcgtcagc----- 5364
|

QY 77 -agatcgcgtctgctacaaagccgaaagaaagctgagttgctgcgtcgcacgcgtgagc 135
|||||
DB 5365 gagatccgcgtctgctacaaagccgaaagaaagctgagttgctgcgtcgcacgcgtgagc 5424
|||||

QY 136 aataactag 144
|||||
DB 5425 aataactag 5433
|||||

RESULT 3
A36847 717 bp DNA PAT 05-MAR-1997
LOCUS Sequence 1 from Patent WO9402510.
DEFINITION A36847
ACCESSION A36847
VERSION A36847.1 GI:2294105
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 717)
AUTHORS Bauer, W., Breckenridge, R., Cardinaux, F., Gombert, F., Gram, H.,
Ramage, P., Schneider, H., Waelchli, R., Albert, Rainer and Lewis, I.
TITLE ANALOGS OF PTH
JOURNAL Patent: WO 9402510-A 1 03-FEB-1994;
SANDOZ AG (AT)

COMMENT Other publication HU 70459 951030
Other publication CZ 9500088 951018
Other publication AU 4156693 940120
Other publication SK 4395 950607
Other publication NZ 248137 951221
Other publication GB 2269176 940202
Other publication JP 6184198 940705
Other publication CA 2100423 940116
Other publication CN 1099801 950308
Other publication DE 43933817 950427
Other publication NO 950123 950315
Other publication ZA 9305126 950116
Other publication FI 950171 950313.

FEATURES
source 1..717
Location/Qualifiers
/organism="unidentified"
/db_xref="taxon:32644"
101..562
/note="unnamed protein product"
/codon_start=1
/protein_id="CAA02267.1"
/db_xref="GI:2294106"
/translation="MSKPKRYVYNNKELLOAIIDDKTELANKDPKVVYRNDTIG
LAIMLIAGLSKRENFSGYTOSMKOEMADIEASIKGLHNFDETKKYPATITTOAC
FNATVQRSIDPVSSEITQIMHNHLSKINSEKRVEMLRKKLDVHNEVALG"

BASE COUNT 245 a 137 c 144 g 191 t
ORIGIN

Query Match 52.1%; Score 75; DB 6; Length 717;
Best Local Similarity 100.0%; Pred. No. 3.2e-13;
Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 70 ctgcagcaatcgcgtctgctacaaagccgaaagaaagctgagttgctgcgtcgcacgc 129
|||||
DB 563 ctgcagcaatcgcgtctgctacaaagccgaaagaaagctgagttgctgcgtcgcacgc 622
|||||

QY 130 ctgcagcaatcgcgtctgctacaaagccgaaagaaagctgagttgctgcgtcgcacgc 144
|||||
DB 623 ctgcagcaatcgcgtctgctacaaagccgaaagaaagctgagttgctgcgtcgcacgc 637
|||||

RESULT 4
A36849 945 bp DNA PAT 05-MAR-1997
LOCUS Sequence 3 from Patent WO9402510.
DEFINITION A36849
ACCESSION A36849
VERSION A36849.1 GI:2294107
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 945)
AUTHORS Bauer, W., Breckenridge, R., Cardinaux, F., Gombert, F., Gram, H.,
Ramage, P., Schneider, H., Waelchli, R., Albert, Rainer and Lewis, I.
TITLE ANALOGS OF PTH
JOURNAL Patent: WO 9402510-A 3 03-FEB-1994;
SANDOZ AG (AT)

COMMENT Other publication HU 70459 951030
Other publication CZ 9500088 951018
Other publication AU 4156693 940120
Other publication SK 4395 950607
Other publication NZ 248137 951221
Other publication GB 2269176 940202
Other publication JP 6184198 940705
Other publication CA 2100423 940116
Other publication CN 1099801 950308
Other publication DE 43933817 950427
Other publication NO 950123 950315
Other publication ZA 9305126 950116
Other publication FI 950171 950313.

FEATURES
source 1..945
Location/Qualifiers

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 6, 2002, 09:13:46 ; Search time 71.9 Seconds
(without alignments)
2.191 Million cell updates/sec

Title: US-09-897-042-14

Perfect score: 37

Sequence: 1 MNLVPMN 7

Scoring table:

BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database : Issued_Patents_AA: *
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep: *
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep: *
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep: *
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep: *
5: /cgn2_6/ptodata/2/1aa/PCTUS.COMB.pep: *
6: /cgn2_6/ptodata/2/1aa/backfiles1.pep: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	31	83.8	2071	4 US-09-415-522-6	Sequence 6, Appli
2	29	78.4	237	3 US-08-924-747-28	Sequence 28, Appli
3	29	78.4	237	4 US-09-247-373B-28	Sequence 28, Appli
4	29	78.4	237	4 US-09-236-715-28	Sequence 28, Appli
5	28	75.7	75	1 US-08-428-415-14	Sequence 14, Appli
6	28	75.7	75	1 US-08-379-685-14	Sequence 14, Appli
7	28	75.7	75	2 US-08-854-029-14	Sequence 14, Appli
8	28	75.7	75	4 US-08-428-762-14	Sequence 14, Appli
9	28	75.7	211	1 US-08-428-415-9	Sequence 9, Appli
10	28	75.7	211	1 US-08-379-685-9	Sequence 9, Appli
11	28	75.7	211	2 US-08-854-029-9	Sequence 9, Appli
12	28	75.7	211	4 US-08-428-762-9	Sequence 9, Appli
13	28	75.7	396	3 US-09-082-310-2	Sequence 2, Appli
14	28	75.7	2500	2 US-08-801-263A-2	Sequence 2, Appli
15	28	75.7	2500	2 US-09-102-248-2	Sequence 2, Appli
16	28	75.7	2512	2 US-08-801-263A-9	Sequence 9, Appli
17	28	75.7	2512	2 US-09-102-248-9	Sequence 9, Appli
18	28	75.7	2517	3 US-08-801-263A-5	Sequence 5, Appli
19	28	75.7	2517	3 US-09-102-248-5	Sequence 5, Appli
20	27	73.0	214	4 US-09-587-066-6	Sequence 6, Appli
21	27	73.0	273	1 US-08-118-270-63	Sequence 63, Appli
22	27	73.0	273	5 PCT-US93-08528-63	Sequence 63, Appli
23	27	73.0	398	1 US-08-446-777-2	Sequence 2, Appli
24	27	73.0	671	6 5266464-2	Sequence 2, Appli
25	27	73.0	694	2 US-08-701-240-2	Sequence 2, Appli
26	27	73.0	694	4 US-09-138-236-2	Sequence 2, Appli
27	27	73.0	695	2 US-08-701-240-4	Sequence 4, Appli

28	27	73.0	695	4 US-09-138-236-4	Sequence 4, Appli
29	27	73.0	767	1 US-08-446-777-6	Sequence 6, Appli
30	27	73.0	767	1 US-08-446-777-8	Sequence 8, Appli
31	27	73.0	1012	1 US-08-219-262B-10	Sequence 10, Appli
32	27	73.0	1012	3 US-09-031-655-10	Sequence 10, Appli
33	27	73.0	1013	2 US-08-708-541A-34	Sequence 34, Appli
34	27	73.0	1824	2 US-08-680-327-3	Sequence 3, Appli
35	27	73.0	1824	4 US-09-228-246-2	Sequence 2, Appli
36	26	70.3	9	3 US-09-075-257A-1	Sequence 1, Appli
37	26	70.3	9	4 US-09-534-639-1	Sequence 1, Appli
38	26	70.3	10	3 US-09-075-257A-10	Sequence 10, Appli
39	26	70.3	10	4 US-09-534-639-10	Sequence 10, Appli
40	26	70.3	15	4 US-08-256-104-5	Sequence 5, Appli
41	26	70.3	15	4 US-08-408-930A-4	Sequence 4, Appli
42	26	70.3	25	3 US-09-075-257A-17	Sequence 17, Appli
43	26	70.3	25	4 US-09-534-639-17	Sequence 17, Appli
44	26	70.3	28	3 US-09-075-257A-15	Sequence 15, Appli
45	26	70.3	28	3 US-09-075-257A-16	Sequence 16, Appli
46	26	70.3	28	4 US-09-534-639-15	Sequence 15, Appli
47	26	70.3	28	4 US-09-534-639-16	Sequence 16, Appli
48	26	70.3	29	3 US-09-075-257A-13	Sequence 13, Appli
49	26	70.3	29	3 US-09-075-257A-14	Sequence 14, Appli
50	26	70.3	29	3 US-09-075-257A-18	Sequence 18, Appli

ALIGNMENTS

```
RESULT 1
US-09-415-522-6
Sequence 6, Application US/09415522A
Patent No. 6291660
GENERAL INFORMATION:
APPLICANT: Gaffney, Thomas
APPLICANT: Mendland, Juergen
APPLICANT: Philippsen, Peter
TITLE OF INVENTION: Development
FILE REFERENCE: CGC2046
CURRENT APPLICATION NUMBER: US/09/415,522A
CURRENT FILING DATE: 1998-10-08
NUMBER OF SEQ ID NOS: 28
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 6
LENGTH: 2071
TYPE: PRT
ORGANISM: Ashbya gossypii
US-09-415-522-6

Query Match      83.8%: Score 31; DB 4; Length 2071;
Best Local Similarity 85.7%: Pred. No. 1.9e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 MNLVPMN 7
      ||| |||
Db      718 MNLVPMN 724

RESULT 2
US-08-924-747-28
Sequence 28, Application US/08924747
Patent No. 6063570
GENERAL INFORMATION:
APPLICANT: MCGONIGLE, BRIAN
APPLICANT: O'KEEFE, DANIEL
TITLE OF INVENTION: SOYBEAN GLUTATHIONE-S-TRANSFERASE
TITLE OF INVENTION: ENZYMES
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESSES:
ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
STREET: 1007 MARKET STREET
CITY: WILMINGTON
```

STATE: DELAWARE
COUNTRY: UNITED STATES OF AMERICA
ZIP: 19898
COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE, 3.50 INCH
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: MICROSOFT WORD FOR WINDOWS 95
SOFTWARE: MICROSOFT WORD VERSION 7.0A
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/924,747
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: FLOYD, LINDA AXAMETHY
REGISTRATION NUMBER: 33,692
REFERENCE/DOCKET NUMBER: CL-1108
TELECOMMUNICATION INFORMATION:
TELEPHONE: 302-892-8112
TELEFAX: 302-773-0164
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 237 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: protein
ORIGINAL SOURCE:
TISSUE TYPE: SOYBEAN
IMMEDIATE SOURCE:
CLONE: SE1.PK0017.F5
US-08-924-747-28

Query Match 78.4%; Score 29; DB 3; Length 237;
Best Local Similarity 71.4%; Pred. No. 50;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 MNLVPMN 7
:||||:|
Db 55 INLVPMN 61

RESULT 3
US-09-247-373B-28
Sequence 28, Application US/09247373B
Patent No. 6168954
GENERAL INFORMATION:
APPLICANT: MCGONIGLE, BRIAN
APPLICANT: O'KEEFE, DANIEL
TITLE OF INVENTION: SOYBEAN GLUTATHIONE-S-TRANSFERASE ENZYMES
FILE REFERENCE: CL-1108-A
CURRENT APPLICATION NUMBER: US/09/247,373B
CURRENT FILING DATE: 1999-02-10
PRIOR APPLICATION NUMBER: 08/924,747
PRIOR FILING DATE: 1997-09-05
NUMBER OF SEQ ID NOS: 56
SOFTWARE: Microsoft Office 97
SEQ ID NO 28
LENGTH: 237
TYPE: PRT
ORGANISM: SOYBEAN
US-09-247-373B-28

Query Match 78.4%; Score 29; DB 4; Length 237;
Best Local Similarity 71.4%; Pred. No. 50;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 MNLVPMN 7
:||||:|
Db 55 INLVPMN 61

RESULT 4
US-09-296-715-28
Sequence 28, Application US/09296715
Patent No. 6171839
GENERAL INFORMATION:
APPLICANT: MCGONIGLE, BRIAN
APPLICANT: O'KEEFE, DANIEL
TITLE OF INVENTION: SOYBEAN GLUTATHIONE-S-TRANSFERASE
TITLE OF INVENTION: ENZYMES
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: E.I. DU PONT DE NEMOURS AND COMPANY
STREET: 1007 MARKET STREET
CITY: WILMINGTON
STATE: DELAWARE
COUNTRY: UNITED STATES OF AMERICA
ZIP: 19898
COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE, 3.50 INCH
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: MICROSOFT WORD FOR WINDOWS 95
SOFTWARE: MICROSOFT WORD VERSION 7.0A
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/296,715
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: FLOYD, LINDA AXAMETHY
REGISTRATION NUMBER: 33,692
REFERENCE/DOCKET NUMBER: CL-1108
TELECOMMUNICATION INFORMATION:
TELEPHONE: 302-892-8112
TELEFAX: 302-773-0164
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 237 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: protein
ORIGINAL SOURCE:
TISSUE TYPE: SOYBEAN
IMMEDIATE SOURCE:
CLONE: SE1.PK0017.F5
US-09-296-715-28

Query Match 78.4%; Score 29; DB 4; Length 237;
Best Local Similarity 71.4%; Pred. No. 50;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 MNLVPMN 7
:||||:|
Db 55 INLVPMN 61

RESULT 5
US-08-428-415-14
Sequence 14, Application US/08428415
Patent No. 5756335
GENERAL INFORMATION:
APPLICANT: Cold Spring Harbor Laboratory
TITLE OF INVENTION: No. 5756335el Human cdc25 Genes, Encoded
TITLE OF INVENTION: Products and Uses Thereof
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02109
COMPUTER READABLE FORM:

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 6, 2002, 09:13:46 ; Search time 103.34 Seconds
(without alignments)
5.018 Million cell updates/sec

Title: US-09-897-042-14

Sequence: 1 MNLVPMN 7

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 50 summaries

Database :

A_Geneseq_1101.*
1: /SIDS2/gcgdata/geneseq/geneseq/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseq/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseq/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseq/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseq/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseq/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseq/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseq/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseq/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseq/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseq/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseq/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseq/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseq/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseq/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseq/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseq/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseq/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseq/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	31	83.8	2071	21	AAV84686
2	29	78.4	237	21	AAV84686
3	29	78.4	237	21	AAV84686
4	29	78.4	237	21	AAV84686
5	29	78.4	237	21	AAV84686
6	29	78.4	237	21	AAV84686
7	29	78.4	237	21	AAV84686
8	29	78.4	237	21	AAV84686
9	29	78.4	237	21	AAV84686
10	29	78.4	237	21	AAV84686
11	29	78.4	237	21	AAV84686

12	28	75.7	48	20	AAV14446
13	28	75.7	101	18	AAV16679
14	28	75.7	215	21	AAV16004
15	28	75.7	344	22	AAV39369
16	28	75.7	353	22	AAV41155
17	28	75.7	364	22	AAV83291
18	28	75.7	391	21	AAV4612
19	28	75.7	391	21	AAV94155
20	28	75.7	396	21	AAV44277
21	28	75.7	396	22	AAV38840
22	28	75.7	407	22	AAV40626
23	28	75.7	407	19	AAV70461
24	28	75.7	807	19	AAV70463
25	28	75.7	807	19	AAV70465
26	27	73.0	48	22	AAV33402
27	27	73.0	56	22	AAV59287
28	27	73.0	110	22	AAV5970
29	27	73.0	115	22	AAV5975
30	27	73.0	128	22	AAV5972
31	27	73.0	136	21	AAV01483
32	27	73.0	170	21	AAV96467
33	27	73.0	214	22	AAV00712
34	27	73.0	216	22	AAV98603
35	27	73.0	216	22	AAV98606
36	27	73.0	248	22	AAV87636
37	27	73.0	273	15	AAV48742
38	27	73.0	273	17	AAV02714
39	27	73.0	275	22	AAV95047
40	27	73.0	278	21	AAV36698
41	27	73.0	278	21	AAV37341
42	27	73.0	278	21	AAV43322
43	27	73.0	279	21	AAV42870
44	27	73.0	317	13	AAV27870
45	27	73.0	317	11	AAV3007
46	27	73.0	331	11	AAV05934
47	27	73.0	365	22	AAV01864
48	27	73.0	372	21	AAV36697
49	27	73.0	372	21	AAV37340
50	27	73.0	372	21	AAV43321

ALIGNMENTS

RESULT 1	
ID	AAV84686 standard; Protein: 2071 AA.
XX	AAV84686;
DT	08-AUG-2000 (first entry)
XX	
DE	Amino acid sequence of AG003, a GTPase activating protein.
XX	
KW	GTP-binding protein; AG003; GTPase activating protein; fungal growth;
KW	phosphatidylinositol-4 kinase; cytokinesis gene; fungal development;
KW	fungicidal; filamentous fungi; plant pathogen; Septoria tritici;
KW	Stagnospora nodorum; Magnaporthe grisea; human pathogen;
KW	Candida albicans; Aspergillus fumigatus.
XX	
OS	Asbya gossypii.
XX	
PN	WO200022133-A1.
XX	
PD	20-APR-2000.
XX	
PF	06-OCT-1999; 99WO-EP07501.
XX	
PR	08-OCT-1998; 98US-0168804.
XX	
PA	(NOVS) NOVARTIS AG.
XX	(NOVS) NOVARTIS-ERFINDUNGEN VERW GES MBH.

PI Gaffney TD, Wendland J, Dietrich F, Philippson P, Goff SA:
XX WPI: 2000-317998/27.
DR N-PSDB; AAA14503.
XX
PT Ashbya gossypii nucleotide and protein sequences useful for identifying
PT fungal growth inhibitors
XX
PS Claim 2; Page 48-60; 93pp; English.
XX
CC The present sequence represents a putative GTPase activating protein,
CC designated AG003. The specification also describes putative GTP
CC binding protein genes, putative phosphatidylinositol-4 kinase
CC protein gene, and a putative cytokinesis gene. These genes are
CC essential for fungal growth and development. The proteins can be used
CC in methods to identify compounds that have fungicidal activity. Compounds
CC with fungicidal activity can be used for suppressing fungal growth,
CC especially of filamentous fungi. Fungi that can be suppressed include
CC plant pathogens (e.g. Septoria tritici, Stagonospora nodorum, and
CC Magnaporthe grisea) and human pathogens (e.g. Candida albicans, and
CC Aspergillus fumigatus).
XX
SQ Sequence 2071 AA:

Query Match 83.8%; Score 31; DB 21; Length 2071;
Best Local Similarity 85.7%; Pred. No. 3e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 MNLVPMN 7
Db 718 mmlppm 724

RESULT 2
AAB07839
ID AAB07839 standard; Protein; 237 AA.
XX
AC AAB07839;
XX
DT 14-NOV-2000 (first entry)
XX
DE Amino acid sequence of a soybean type IV glutathione-S-transferase.
XX
KM Soybean; glutathione-S-transferase; GST; detoxification;
KM xenobiotic compound; herbicide-tolerance; transgenic plant;
KM herbicide synergist.
XX
OS Glycine max.
XX
PN WO200047728-A2.
XX
PD 17-AUG-2000.
XX
PE 10-FEB-2000; 2000WO-US03347.
XX
PR 10-FEB-1999; 99US-0247373.
XX
PA (DUPO) DU PONT DE NEMOURS & CO E I.
XX
PI McConigle B, O'Keefe DP;
XX
DR WPI: 2000-549144/50.
DR N-PSDB; AAA54470.
XX
PT Soybean glutathione-S-transferase polypeptides and polynucleotides used
PT to produce herbicide tolerant transgenic plants and to screen for
PT inhibitors or substrates of the enzyme -
XX
PS Claim 4; Page 69-70; 84pp; English.
XX
CC The present sequence represents a soybean glutathione-S-transferase
CC protein. The enzyme is involved in the detoxification of

CC xenobiotic compounds in plants and seeds. The GST polynucleotides
CC and polypeptides are used for the production of herbicide-tolerant
CC transgenic plants, and for the development of screening assays to
CC identify GST inhibitors and substrates, which can be used as
CC herbicide synergists. GST gene specific probes can be used in gene
CC identification methods. The recombinant GST enzymes can be used to
CC produce enzyme specific antibodies which are used to detect the
CC enzymes in situ in cells or in vitro in cell extracts.
XX
SQ Sequence 237 AA:

Query Match 78.4%; Score 29; DB 21; Length 237;
Best Local Similarity 71.4%; Pred. No. 79;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
OY 1 MNLVPMN 7
Db 55 mlvpin 61

RESULT 3
AAB03744
ID AAB03744 standard; Protein; 237 AA.
XX
AC AAB03744;
XX
DT 04-OCT-2000 (first entry)
XX
DE Clone sel.pK0017.f5 type IV GST protein sequence.
XX
KM Soybean; glutathione-S-transferase; GST; detoxify; herbicide; stress;
KM transgenic plant; tolerant; plant breeding.
XX
OS Glycine max.
XX
PN US6063570-A.
XX
PD 16-MAY-2000.
XX
PE 05-SEP-1997; 97US-0924747.
XX
PR 05-SEP-1997; 97US-0924747.
XX
PA (DUPO) DU PONT DE NEMOURS & CO E I.
XX
PI McConigle B, O'Keefe DP;
XX
DR WPI: 2000-375487/32.
DR N-PSDB; AAA53406.
XX

XX New Glutathione-S-transferase enzymes and isolated nucleic acid
XX fragments encoding them, useful for detoxifying xenobiotic compounds in
XX plants and seeds, as well as in producing transgenic plants that are
XX herbicide-resistant -
XX
PS Claim 1; Column 65-66; 36pp; English.
XX
CC This sequence represents a Glutathione-S-Transferase (GST) protein
CC isolated from a soybean clone. The invention relates to isolated nucleic
CC acid fragments (see AAA53393-A53406) which encode soybean GST
CC polypeptides (AAB03731-B03744). GSTs are a family of enzymes which
CC catalyse the conjugation of glutathione, homogluthathione and other
CC electrophilic compounds, to a large range of hydrophobic,
CC of certain herbicides. The GST nucleotide sequences are useful in the
CC construction of herbicide-tolerant transgenic plants, plants that are
CC tolerant to a wide variety of stresses, or plants in which the GST
CC enzymes are present at higher or lower levels than they are normally. The
CC nucleic acid fragments are also useful as probes for genetically and
CC physically mapping the genes that they are part of, and as markers for
CC traits linked to expression of the enzymes. This will be useful in plant
CC breeding in order to develop lines with desired phenotypes or in the

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 6, 2002, 09:15:21 ; Search time 135.97 Seconds
(without alignments)
7.530 Million cell updates/sec

Title: US-09-897-042-14

Perfect score: 37

Sequence: 1 MNLVPMN 7

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 50 summaries

Database :

1: SPREMBL_17:*
2: sp_archaea:*
3: sp_bacteria:*
4: sp_fungi:*
5: sp_human:*
6: sp_invertebrate:*
7: sp_mmc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	33	89.2	2451	12	Q9WJD4
2	32	86.5	336	2	Q9EVAL
3	32	86.5	346	2	Q9RVT6
4	32	86.5	492	5	Q27470
5	32	86.5	561	2	Q9K6M4
6	31	83.8	407	10	Q9L7X6
7	31	83.8	407	10	Q82118
8	31	83.8	1037	5	Q9NHF7
9	31	83.8	2071	3	Q9HF75
10	30	81.1	346	2	Q91206
11	30	81.1	369	2	Q30580
12	30	81.1	619	10	Q9Z084
13	30	81.1	684	4	Q9UC31
14	30	81.1	781	10	Q9Z0Q3
15	30	81.1	827	2	Q31506
16	30	81.1	1431	10	Q9C6V0
17	30	81.1	2325	5	Q18104
18	29	81.1	2475	12	Q08358
19	29	78.4	113	12	Q66234

20	29	78.4	128	12	Q65288	065288 human adeno
21	29	78.4	237	10	Q9FQ95	Q9FQ95 glycine max
22	29	78.4	280	5	Q9V107	Q9V107 drosophila
23	29	78.4	325	12	Q9QEE7	Q9QEE7 indian citr
24	29	78.4	368	2	Q86754	Q86754 streptomyc
25	29	78.4	408	5	P91470	P91470 caenorhabd
26	29	78.4	424	10	Q9Z0Y3	Q9Z0Y3 arabidopsi
27	29	78.4	496	5	Q16482	Q16482 caenorhabd
28	29	78.4	509	5	Q27482	Q27482 caenorhabd
29	29	78.4	509	5	Q27499	Q27499 caenorhabd
30	29	78.4	723	2	Q9AM59	Q9AM59 actinobact
31	29	78.4	732	5	Q17619	Q17619 caenorhabd
32	29	78.4	933	10	Q9FPR3	Q9FPR3 arabidopsi
33	29	78.4	1879	12	Q9WJ11	Q9WJ11 venezuelan
34	29	78.4	2447	12	Q9WJ10	Q9WJ10 venezuelan
35	29	78.4	2455	12	Q9WJ08	Q9WJ08 venezuelan
36	29	78.4	2477	12	Q9WJ01	Q9WJ01 venezuelan
37	29	78.4	2478	12	Q9WJ06	Q9WJ06 venezuelan
38	29	78.4	2485	12	Q9WJ05	Q9WJ05 venezuelan
39	29	78.4	2485	12	Q9JF75	Q9JF75 venezuelan
40	29	78.4	2485	12	Q9IR88	Q9IR88 venezuelan
41	29	78.4	2485	12	Q9ICG5	Q9ICG5 venezuelan
42	29	78.4	2492	12	Q9IR87	Q9IR87 venezuelan
43	29	78.4	2493	12	Q66594	Q66594 venezuelan
44	29	78.4	2493	12	Q90163	Q90163 venezuelan
45	29	78.4	2493	12	Q66592	Q66592 venezuelan
46	29	78.4	2497	12	Q88978	Q88978 venezuelan
47	29	78.4	2498	12	Q86924	Q86924 aura virus
48	29	78.4	2499	12	Q9WJ07	Q9WJ07 venezuelan
49	29	78.4	2501	12	Q9WJ09	Q9WJ09 venezuelan
50	29	78.4	2651	10	Q9FR85	Q9FR85 arabidopsi

ALIGNMENTS

RESULT 1	ID	Q9WJD4	PRELIMINARY:	PRT: 2451 AA.
AC	Q9WJD4:			
DT	01-NOV-1999	(TREMBLrel. 12, Created)		
DT	01-NOV-1999	(TREMBLrel. 12, Last sequence update)		
DT	01-JUN-2001	(TREMBLrel. 17, Last annotation update)		
DE	NONSTRUCTURAL POLYPEPTIDE.			
OS	Venezuelan equine encephalitis virus.			
OC	Viruses: ssRNA positive-strand viruses, no DNA stage; Togaviridae;			
OX	Alphavirus.			
RN	NCBI_TaxID=11036;			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-CABASSOU CAAR 508;			
RX	MEDLINE=99101297; PubMed=9886206;			
RA	kinney R.M., Pfeiffer M., Tsuchiya K.R., Chang G.J., Roehrig J.T.;			
RT	Nucleotide sequences of the 26S mRNAs of the viruses defining the			
RL	Venezuelan equine encephalitis antigenic complex.";			
AM	Am. J. Trop. Med. Hyg. 59:952-964(1998).			
[2]	SEQUENCE FROM N.A.			
RC	STRAIN-CABASSOU CAAR 508;			
RA	kinney R.M., Pfeiffer M., Weissenner J.;			
RL	Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.			
DR	EMBL: AF075259; AADI4566.1; -			
DR	InterPro: IPR000606; AADI4566.1; -			
DR	InterPro: IPR002589; DUF27			
DR	InterPro: IPR002520; Peptidase_C9.			
DR	InterPro: IPR01443; Viral_helicase1; 1.			
DR	Pfam: PF01661; DUF27; 1.			
DR	Pfam: PF01707; Peptidase_C9; 1.			
KW	SMART: SM00506; A1pp; 1.			
FT	CHAIN	1	535	METHYLTRANSFERASE NSP1.
FT	CHAIN	536	1329	NONSTRUCTURAL PROTEINASE NSP2.
FT	CHAIN	1330	1844	NONSTRUCTURAL PROTEIN NSP3.

FT CHAIN 1845 2451 RNA POLYMERASE NSP4.
SQ SEQUENCE 2451 AA; 272975 MW; 2253DAEBF24F0383 CRC64;

Query Match
Best Local Similarity 85.7%; Score 33; DB 12; Length 2451;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 MNLVPMN 7
DB 1132 INLVPMN 1138

RESULT 2

O9EVAL PRELIMINARY; PRT; 336 AA.

AC O9EVAL;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CYTOTOXIN K.
GN CYTK.
OS Bacillus cereus.
OC Bacillus; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1396;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed-11069652;
RA Lund T., De Buyser M.L., Granum P.;
RT "A new cytotoxin from Bacillus cereus that may cause necrotic enteritis."
RL EMBL: AJ277962; CAC08441.1;
DR InterPro: IPR001340; Hemlign_pore.
SQ SEQUENCE 336 AA; 36953 MW; 54C1148EAF1B709E CRC64;

Query Match
Best Local Similarity 100.0%; Score 32; DB 2; Length 336;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 NLVPMN 7
DB 248 NLVPMN 253

RESULT 3

O9RV76 PRELIMINARY; PRT; 346 AA.

AC O9RV76;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE CONSERVED HYPOTHETICAL PROTEIN.
GN DR0936.
OS Deinococcus radiodurans.
OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RI;
RC MEDLINE=20036896; PubMed-10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.S., Lam P., McDonald L., Uitterlind T., Zalewski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus radiodurans RI."
RL Science 286:1571-1577(1999).

DR EMBL: AE001946; AAF10513.1; -.
DR TIGR: DR0936; -.
KW Complete proteome.
SQ SEQUENCE 346 AA; 37979 MW; 64506136A45F9C43 CRC64;

Query Match
Best Local Similarity 71.4%; Score 32; DB 2; Length 346;

Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 MNLVPMN 7
DB 279 VNLVPMN 285

RESULT 4

O27470 PRELIMINARY; PRT; 492 AA.

AC O27470;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE SIMILAR TO CYTOCHROME P450 (EC 1.14.14.1).
GN C12D5.7.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilita; Rhabdilitoidea;
OC Rhabdilitidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=94150718; PubMed-7906398;
RA Wilson R., Alnsough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnson L.,
RA Jones M., Kershaw J., Kirsten J., Lalster N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA Smaildon N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.; III of C.
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans."
RL Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Miller N., Stellyes L., Bradshaw H.;
RL Submitted (MAY-1996) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Waterston R.;
RL Submitted (APR-1996) to the EMBL/Genbank/DBJ databases.
CC -i- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL: U55365; AAA98571.1; -.
DR InterPro: IPR001128; CYL_P450.
DR Pfam: PF00067; P450; 1.
DR PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_1.
KW Electron transport; Heme; Membrane; Monooxygenase; Oxidoreductase.
FT BINDING 436 436 HEME (BY SIMILARITY).
SQ SEQUENCE 492 AA; 56770 MW; 52B71FE98FB06C4 CRC64;

Query Match
Best Local Similarity 86.5%; Score 32; DB 5; Length 492;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 NLVPMN 7
DB 361 NLVPMN 366

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 6, 2002, 09:16:56 : Search time 68.88 Seconds
(without alignments)
3.726 Million cell updates/sec

Title: US-09-897-042-14

Perfect score: 37

Sequence: 1 MNLVPMN 7

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 3664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 50 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
1	31	83.8	210 1	CLPP_CANCER
2	29	78.4	142 1	SECB_BUCAI
3	29	78.4	364 1	YS80_MYCTU
4	29	78.4	575 1	P07275_YEAST
5	29	78.4	679 1	PAN3_YEAST
6	29	78.4	2485 1	POLN_EEYV3
7	29	78.4	2492 1	POLN_EEYV3
8	29	78.4	2492 1	POLN_EEYV3
9	29	78.4	2492 1	POLN_EEYV3
10	28	75.7	211 1	LIPB_BUCAI
11	28	75.7	215 1	FIMI_ECOLI
12	28	75.7	317 1	Y402_RICPR
13	28	75.7	363 1	Y402_RICPR
14	28	75.7	389 1	NCAP_CVH22
15	28	75.7	535 1	VS50_HSVSA
16	28	75.7	536 1	CCAA_DROME
17	28	75.7	596 1	MP1P_SCHPO
18	28	75.7	617 1	ASMA_ECOLI
19	28	75.7	905 1	GYRA_RICPR
20	28	75.7	2512 1	POLN_SINDV
21	27	73.0	112 1	OLF8_MOUSE
22	27	73.0	187 1	KIP2_HUMAN
23	27	73.0	187 1	KIP2_MOUSE
24	27	73.0	298 1	Y680_HAEIN
25	27	73.0	317 1	OLF2_RAT
26	27	73.0	345 1	TOB2_MOUSE
27	27	73.0	379 1	Y2A6_PSEAE
28	27	73.0	395 1	SYW_AQDAE
29	27	73.0	400 1	BAG1_HUMAN
30	27	73.0	402 1	BAG1_BOVIN
31	27	73.0	406 1	YMX2_CAEEL
32	27	73.0	472 1	YWIE_CAEEL
33	27	73.0	493 1	CPEI_MOUSE

34	27	73.0	493 1	CPEI_RAT	P05182_rattus norv
35	27	73.0	501 1	CPJ5_MOUSE	O54749_mus musculu
36	27	73.0	505 1	CAT4_METBA	O03662_methanosarc
37	27	73.0	533 1	YADG_SCHPO	O09837_schizosacch
38	27	73.0	553 1	5MTD_VIBCH	O09430_vibrio chol
39	27	73.0	569 1	YES9_MYCPN	P75324_mycoplasma
40	27	73.0	671 1	KPC1_HUMAN	P05771_homo sapien
41	27	73.0	671 1	KPC1_RABIT	P05772_oryctolagus
42	27	73.0	671 1	KPC1_RAT	P04410_rattus norv
43	27	73.0	673 1	KPC2_BOVIN	P05126_bos taurus
44	27	73.0	673 1	KPC2_HUMAN	P05127_homo sapien
45	27	73.0	673 1	KPC2_MOUSE	P04411_mus musculu
46	27	73.0	673 1	KPC2_RABIT	P05773_oryctolagus
47	27	73.0	688 1	ENV_MMTVB	P10259_mouse mamma
48	27	73.0	1012 1	POLG_IBDOV	P27276_avian infec
49	26	70.3	91 1	YURS_BACSU	O32160_bacillus su
50	26	70.3	119 1	Y167_UREPA	O09466_ureaplasma

ALIGNMENTS

```

RESULT 1
ID CLPP_CANCER STANDARD; PRT: 210 AA.
AC 087706;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ATP-DEPENDENT CLP PROTEASE PROTEOLYTIC SUBUNIT (EC 3.4.21.92)
DN (ENDOPEPTIDASE CLP).
OS CLPP OR CC1963.
OC Caulobacter crescentus.
OC Bacteria: Proteobacteria; alpha subdivision: Caulobacter group:
OC Caulobacter.
ON NCBI_TaxID=69394;
OM [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CB15N / NA1000;
RX MEDLINE=98429489; PubMed=9755166;
RA Jenal U., Fuchs T.;
  "An essential protease involved in bacterial cell-cycle control.";
  EMBO J. 17:5658-5669(1998).
  [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
  Eisen J., Heidelberg J.F., Alley M.K.R., Ohta N., Maddock J.R.,
  Porocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
  Debroy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
  Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty K.,
  Ueberback T., Tran K., Wolf A., Yamahyevan J., Ermlaeva M., White O.,
  Salzberg S.L., Shapiro L., Venter J.C., Fraser C.M.;
  "Complete genome sequence of Caulobacter crescentus.";
  Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
  - FUNCTION: CLP CLEAVES PEPTIDES IN VARIOUS PROTEINS IN A PROCESS
    THAT REQUIRES ATP HYDROLYSIS. CLP MAY BE RESPONSIBLE FOR A FAIRLY
    GENERAL AND CENTRAL HOUSEKEEPING FUNCTION RATHER THAN FOR THE
    DEGRADATION OF SPECIFIC SUBSTRATES (BY SIMILARITY).
  - CATALYTIC ACTIVITY: HYDROLYSIS OF PROTEINS TO SMALL PEPTIDES IN
    THE PRESENCE OF ATP AND MAGNESIUM. ALPHA-CASEIN IS THE USUAL TEST
    SUBSTRATE. IN THE ABSENCE OF ATP, ONLY OLIGOPEPTIDES SHORTER THAN
    FIVE RESIDUES ARE CLEAVED (SUCH AS SUCCINYL-LEU-TYR-1-NHMEC; AND
    LEU-TYR-LEU-1-TYR-TRP, IN WHICH THE CLEAVAGE OF THE -TYR-1-LEU-
    AND -TYR-1-TRP-BOND ALSO OCCURS).
  - SIMILARITY: BELONGS TO PEPTIDASE FAMILY S14; ALSO KNOWN AS CLPP
    FAMILY.
  -----
  THIS SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way

```

modified and this statement is not removed. Usage by and for commercial
entitles requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL: AJ010321; CA09090.1; -
DR EMBL: AE005869; AAK23938.1; -
DR MEMPRO: S14.003; -
DR TIGR: CC1963; -
DR InterPro: IPR001907; CLP_protease.
DR Pfam: PF00574; CLP_protease; 1.
DR PRINTS: PR00127; CLP_PROTEASE.
DR PROSITE: PS00381; CLP_PROTEASE_SER; 1.
DR PROSITE: PS00382; CLP_PROTEASE_HIS; 1.
DR HydroLase: Serine protease; Complete proteome.
FT ACT_SITE 107 107 PROBABLE.
FT ACT_SITE 132 132 PROBABLE.
SQ SEQUENCE 210 AA; 23120 MW; CC2421D44C1E7E59 CRC64;

Query Match 83.8%; Score 31; DB 1; Length 210;
Best Local Similarity 100.0%; Pred. No. 8.5;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNLVPM 6
DB 10 MNLVPM 15

RESULT 2
SECB_BUCAI STANDARD; PRT; 142 AA.
ID SECB_BUCAI
AC P57161;
DT 20-AUG-2001 (Rel. 40; Created)
DT 20-AUG-2001 (Rel. 40; Last sequence update)
DT 20-AUG-2001 (Rel. 40; Last annotation update)
DE PROTEIN-EXPORT PROTEIN SECB.
GN SECB OR BU053.
OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
symbiotic bacterium).
OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
OX NCBI_TaxID=118099;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TOKYO 1998;
RA MEDLINE=20445173; PubMed=10993077;
RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
RT "Genome sequence of the endocellular bacterial symbiont of aphids
Buchnera sp. APS.";
RL Nature 407:81-86(2000).
CC -1- FUNCTION: THIS IS ONE OF THE PROTEINS REQUIRED FOR THE NORMAL
EXPORT OF ENVELOPE PROTEINS OUT OF THE CELL CYTOPLASM; IT MAY
BE INVOLVED IN THE INITIATION OF THE EXPORTING PROCESS, BY BINDING
TO THE NASCENT POLYPEPTIDE VIA A SIGNAL SEQUENCE, MAINTAINING
A STABLE AND PRE-TRANSLATION CONFORMATION (BY SIMILARITY).
CC -1- SUBUNIT: PART OF THE PROKARYOTIC PROTEIN TRANSLLOCATION APPARATUS
WHICH COMPRISE SECA, SECB, SECE, SECD AND SECY
(BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE SECB FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entitles requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC EMBL: AP001118; BAB12776.1; -
DR InterPro: IPR003708; SECB.
DR Pfam: PF02556; SECB; 1.
KW Protein transport; Translocation; Complete proteome.
SQ SEQUENCE 142 AA; 16571 MW; C81D176607D91D9 CRC64;

Query Match 78.4%; Score 29; DB 1; Length 142;
Best Local Similarity 71.4%; Pred. No. 16;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MNLVPM 7
DB 126 MNLVPM 132

RESULT 3
YS80_MYCTU STANDARD; PRT; 364 AA.
ID YS80_MYCTU
AC Q10806; Q10805;
DT 01-OCT-1996 (Rel. 34; Created)
DT 15-JUL-1998 (Rel. 36; Last sequence update)
DT 20-AUG-2001 (Rel. 40; Last annotation update)
DE HYPOTHETICAL 29.6 KDA PROTEIN RV2879C/RV2879C.
GN RV2880C/RV2879C OR MT2947 OR MTCY274.11C/MTCY274.10C.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Kirogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Uitterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
laboratory strains.";
RL Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE UP063 FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entitles requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC EMBL: Z74024; CA98356.1; ALT_FRAME.
DR EMBL: Z74024; CA98355.1; ALT_FRAME.
DR EMBL: AE007118; AAK47272.1; -
DR TIGR: MT2947; -
DR TubercuList: RV2879C; -
DR TubercuList: RV2880C; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 364 AA; 39235 MW; 74351614D3484A32 CRC64;

Query Match 78.4%; Score 29; DB 1; Length 364;
Best Local Similarity 57.1%; Pred. No. 40;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 6, 2002, 09:13:46 ; Search time 77.97 Seconds
(without alignments)
6.839 Million cell updates/sec

Title: US-09-897-042-14

Perfect score: 37

Sequence: 1 MNLVPMN 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 50 summaries

Database :

1: PIR.68:*
2: PIR1:*
3: PIR2:*
4: PIR3:*
5: PIR4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	32	86.5	346	2 F75457	conserved hypothet
2	32	86.5	492	2 T30066	hypothetical prote
3	32	86.5	561	2 A84113	transposase (12) B
4	31	83.8	407	2 JF0113	zinc-finger protei
5	30	81.1	346	2 A83415	hypothetical prote
6	30	81.1	619	2 A84449	hypothetical prote
7	30	81.1	684	2 T43452	hypothetical prote
8	30	81.1	781	2 T43452	En/Spm-like transp
9	30	81.1	827	2 A69793	conserved hypothet
10	30	81.1	2325	2 T15566	hypothetical prote
11	30	81.1	2475	2 S35307	polypeptide pp220
12	29	78.4	128	2 S33949	early E3B 14.7k pr
13	29	78.4	142	2 H84935	protein-export pro
14	29	78.4	189	2 B70924	conserved hypothet
15	29	78.4	368	2 T35453	hypothetical prote
16	29	78.4	408	2 T29949	hypothetical prote
17	29	78.4	424	2 D84677	hypothetical prote
18	29	78.4	509	2 T29849	hypothetical prote
19	29	78.4	509	2 T29303	hypothetical prote
20	29	78.4	575	1 RDBYC	1-pyrroline-5-carb
21	29	78.4	679	2 S37842	hypothetical prote
22	29	78.4	732	2 T19570	hypothetical prote
23	29	78.4	1015	2 T00726	probable serine/th
24	29	78.4	2492	1 MNMWTB	nonstructural poly
25	29	78.4	2492	1 A44213	nonstructural poly
26	29	78.4	2492	1 C44213	nonstructural poly
27	28	75.7	208	2 A82712	lipoproteinase Clp
28	28	75.7	211	2 B84961	lipoprotein li
29	28	75.7	215	2 S56540	limb protein - Esc

30	28	75.7	215	2 E86129	limbrial protein l
31	28	75.7	220	2 T43857	hypothetical prote
32	28	75.7	317	2 A71698	hypothetical prote
33	28	75.7	330	2 A86355	hypothetical prote
34	28	75.7	343	2 G72218	conserved hypothet
35	28	75.7	362	2 B83963	hypothetical prote
36	28	75.7	363	2 F69878	conserved hypothet
37	28	75.7	389	1 VHIH2E	nucleocapsid prote
38	28	75.7	389	2 S08031	nucleocapsid prote
39	28	75.7	514	2 E86738	multidrug resistanc
40	28	75.7	533	2 F82237	conserved hypothet
41	28	75.7	535	2 C36811	hypothetical prote
42	28	75.7	535	2 A37994	R1 protein - salm
43	28	75.7	596	1 S62407	protein-tyrosine-p
44	28	75.7	617	2 G64972	yegA protein precu
45	28	75.7	905	2 H71731	DNA gyrase chain A
46	28	75.7	2512	1 MNMVS	nonstructural poly
47	28	75.7	2514	1 MNMVS2	nonstructural poly
48	27	73.0	44	2 I37286	olfactory receptor
49	27	73.0	44	2 E37286	hypothetical prote
50	27	73.0	96	2 T18043	

ALIGNMENTS

RESULT 1
F75457
conserved hypothetical protein - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C:Accession: F75457
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.;
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896
A:Accession: F75457
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-346 <WHI>
A:Cross-references: GB:AE001946; GB:AE000513; NID:96458655; PIDN:AAPI0513.1; PID:9645
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR0936
A:Map position: 1
C:Superfamily: conserved hypothetical protein H10365

Query Match 86.5%; Score 32; DB 2; Length 346;
Best Local Similarity 71.4%; Pred. NO. 13;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 MNLVPMN 7
DB 279 VMLIPMN 285

RESULT 2
T30066
hypothetical protein C1205.7 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 16-Feb-2001
C:Accession: T30066
R:Miller, N.; Stellyes, L.; Bradshaw, H.
submitted to the EMBL Data Library, April 1996
A:Description: The sequence of C. elegans cosmid C1205.
A:Reference number: z20729
A:Accession: T30066
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-492 <ML>

A:Cross-references: EMBL:U55365; PIDN:AAA98571.1; GSPDB:GN00023; CESP:C12D5.7
A:Experimental source: strain Bristol N2; clone C12D5
C:Genetics:
A:Gene: CESP:C12D5.7
A:Map position: 5
A:Introns: 28/1: 166/3; 231/2; 354/3
C:Superfamily: human cytochrome P450 CYP2D6; cytochrome P450 homology
C:Keywords: heme; iron; metalloprotein
F:296-458/Domain: cytochrome P450 homology <P45>
F:436/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 86.5%; Score 32; DB 2; Length 492;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2 NMLVPMN 7
b 361 NMLVPMN 366

RESULT 3
AB4113
transposase (12) BH3705 [imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 31-Dec-2000
C:Accession: AB4113
R:Takami, H.; Nakase, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hirai
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A:Reference number: A83650; MUID:20263314
A:Accession: AB4113
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-561 <STO>
A:Cross-references: GB:AP001519; GB:BA000004; NID:g10176109; PIDN:BA07424.1; GSPDB:GN00
A:Experimental source: strain C-125
C:Genetics:
A:Gene: BH3705

Query Match 86.5%; Score 32; DB 2; Length 561;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2 NMLVPMN 7
Db 491 NMLVPMN 496

RESULT 4
JE0113
zinc-finger protein S3574 [imported] - rice
C:Species: Oryza sativa (rice)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 20-Apr-2000
C:Accession: JE0113
R:Song, J.; Yamamoto, K.; Showmura, A.; Itadani, H.; Zhong, H.S.; Yano, M.; Sasaki, T.
DNA Res. 5, 95-101, 1998
A:Title: Isolation and mapping of a family of putative zinc-finger protein cDNAs from ri
A:Reference number: JE0113; MUID:98344140
A:Accession: JE0113
A:Molecule type: mRNA
A:Residues: 1-407 <SON>
A:Cross-references: DDBJ:AB001888; NID:g3618319; PIDN:BAA33206.1; PID:dl034172; PID:g361

Query Match 83.8%; Score 31; DB 2; Length 407;
Best Local Similarity 71.4%; Pred. No. 26;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Oy 1 NMLVPMN 7
Db 177 NMLVPMN 183

RESULT 5
AB3415
hypothetical protein PA1839 [imported] - Pseudomonas aeruginosa (strain PA01)

C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: AB3415
R:Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lapidis, K.; L
Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
A:Reference number: AB2950; MUID:20437337
A:Accession: AB3415
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-346 <STO>
A:Cross-references: GB:AE004610; GB:AE004091; NID:g9947825; PIDN:AG05228.1; GSPDB:GN

A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA1839
C:Superfamily: conserved hypothetical protein H10365

Query Match 81.1%; Score 30; DB 2; Length 346;
Best Local Similarity 71.4%; Pred. No. 37;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 1 NMLVPMN 7
Db 276 NMLVPMN 282

RESULT 6
AB4449
hypothetical protein At2g03480 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: AB4449
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: AB4420; MUID:20083487
A:Accession: AB4449
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-619 <STO>
A:Cross-references: GB:AE002093; NID:g4335751; PIDN:AD17428.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g03480
A:Map position: 2

Query Match 81.1%; Score 30; DB 2; Length 619;
Best Local Similarity 71.4%; Pred. No. 71;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Oy 1 NMLVPMN 7
Db 500 MNVPMN 506

RESULT 7
T43452
hypothetical protein DKFP434E2118.1 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 03-Nov-2000
C:Accession: T43452
R:Gtenwaelder, B.; Obermaier, B.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, December 1999

GenCore version 4.5
Copyright (c) 1993 - 2000 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 6, 2002, 09:48:12 ; Search time 121.65 Seconds
(without alignments)
44,681 Million cell updates/sec

Title: US-09-897-042-13

Perfect score: 24

Sequence: 1 atgaactgtgaccgatgaactaa 24

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/prodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/prodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/prodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/prodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/prodata/2/ina/PCOTUS_COMB.seq:*
6: /cgn2_6/prodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	16.6	69.2	1452	2	US-08-966-316-12
2	16.6	69.2	1452	2	US-08-966-316-12
3	16.4	68.3	2097	6	5460961-4
4	16.4	68.3	2261	4	5460961-1
5	16.2	67.5	658	4	US-09-328-111-816
6	16.2	67.5	1269	1	US-08-265-429A-4
7	16.2	67.5	1269	5	PCT-US95-09069-4
8	16.2	67.5	1401	4	US-08-852-457-1
9	16.2	67.5	3404	1	US-08-265-429A-1
10	16.2	67.5	3404	5	PCT-US95-09069-1
11	16.2	66.7	752	4	US-08-998-416-568
12	16.2	66.7	784	3	US-08-961-083-163
13	16.2	66.7	3135	4	US-09-321-276-1
14	16.2	66.7	3135	4	US-09-321-276-3
15	16.2	66.7	3135	4	US-08-916-481-1
16	16.2	66.7	3135	4	US-08-916-481-1
17	15.6	65.0	1608	1	US-07-621-670-2
18	15.6	65.0	11970	4	US-09-345-217-1
19	15.6	65.0	11970	3	US-09-345-217-1
20	15.6	65.0	152331	3	US-09-128-155-16
21	15.2	63.3	2574	4	US-09-142-529-2
22	15.2	63.3	3828	5	PCT-US93-10500-1
23	15.2	63.3	4190	2	US-08-488-706-3
24	15.2	63.3	4517	5	PCT-US93-06251-83
25	15.2	63.3	7721	3	US-08-772-270A-14
26	15.2	62.5	208	1	US-08-480-552-17
27	15.2	62.5	208	3	US-08-929-208-17

28	15	62.5	1760	3	US-09-082-310-4	Sequence 4, Appl
29	15	62.5	2400	6	5256558-6	Patent No. 5256558
30	15	62.5	3254	1	US-08-162-809-15	Sequence 15, Appl
31	15	62.5	10660	2	US-08-267-803B-8	Sequence 8, Appl
32	15	62.5	10660	4	US-09-041-886-16	Sequence 16, Appl
33	14.8	61.7	1577	2	US-08-717-587A-1	Sequence 1, Appl
34	14.8	61.7	1577	3	US-08-883-610A-1	Sequence 1, Appl
35	14.8	61.7	1577	3	US-08-936-094A-1	Sequence 1, Appl
36	14.8	61.7	4403765	4	US-09-103-840A-2	Sequence 1, Appl
37	14.8	61.7	4411529	4	US-09-103-840A-1	Sequence 1, Appl
38	14.6	60.8	445	1	US-08-592-126-67	Sequence 67, Appl
39	14.6	60.8	467	4	US-08-927-219-133	Sequence 133, Appl
40	14.6	60.8	602	4	US-08-998-416-52	Sequence 52, Appl
41	14.6	60.8	748	4	US-08-998-416-263	Sequence 263, Appl
42	14.6	60.8	751	4	US-08-998-416-399	Sequence 399, Appl
43	14.6	60.8	881	1	US-08-631-200-12	Sequence 12, Appl
44	14.6	60.8	881	1	US-08-829-553-12	Sequence 12, Appl
45	14.6	60.8	881	2	US-08-922-267A-12	Sequence 12, Appl

ALIGNMENTS

RESULT 1
US-08-966-316-12
; Sequence 12, Application US/08966316
; Patent No. 5932445
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Au-Young, Janice
; APPLICANT: Reddy, Roopa
; APPLICANT: Murry, Lynn E.
; APPLICANT: Mathur, Preetle
; TITLE OF INVENTION: SIGNAL PEPTIDE - CONTAINING PROTEINS
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/966,316
; FILING DATE: Herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0424 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1452 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: EOSIHET02
; CLONE: 318060
; :
; US-08-966-316-12

Query Match 69.2%; Score 16.6; DB 2; Length 1452;
Best Local Similarity 82.6%; Pred. No. 32;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 1 atgaactgtgacgatgaacta 23
||| ||||| ||| ||||| |||||
Db 1027 ATGCACCTTCCTACAGATGAAGTGA 1049

RESULT 2
US-08-966-316-12/c
Sequence 12, Application US/08966316
Patent No. 5932445

GENERAL INFORMATION:

APPLICANT: Lal, Preeti
APPLICANT: Au-Young, Janice

APPLICANT: Reddy, Roopa
APPLICANT: Murthy, Lynn E.

APPLICANT: Mathur, Preeti
TITLE OF INVENTION: SIGNAL PEPTIDE - CONTAINING PROTEINS

NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:

ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive

CITY: Palo Alto
STATE: CA

COUNTRY: USA
ZIP: 94304

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/966,316
FILING DATE: Herewith

CLASSIFICATION: 435
PRIOR APPLICATION DATA:

APPLICATION NUMBER:
FILING DATE:

ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.

REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0424 US

TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555

TELEFAX: 650-845-4166
TELEX:

INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:

LENGTH: 1452 base pairs
TYPE: nucleic acid

STRANDEDNESS: single
TOPOLOGY: linear

IMMEDIATE SOURCE:
LIBRARY: EOSIHET02

CLONE: 318060
US-08-966-316-12

Query Match 69.2%; Score 16.6; DB 2; Length 1452;
Best Local Similarity 82.6%; Pred. No. 32;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 1 atgaactgtgacgatgaacta 23
||| ||||| ||| ||||| |||||
Db 410 ATGCACCTTCCTACAGATGAAGTGA 388

RESULT 3
5460961-4/c
Patent No. 5460961

APPLICANT: DEBY, CAROL; PINCEMAIL, JOEL; BOLLEN, ALEX
TITLE OF INVENTION: HUMAN MYELOPEROXIDASE AND ITS
THERAPEUTIC APPLICATION

NUMBER OF SEQUENCES: 13
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/641,678
FILING DATE: 16-JAN-1991

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 460,931

FILING DATE: 14-FEB-1990
SEQ ID NO: 4

LENGTH: 2097
5460961-4

Query Match 68.3%; Score 16.4; DB 6; Length 2097;
Best Local Similarity 94.4%; Pred. No. 41;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 2 tgaactgtgacgatga 19
||||| ||||| ||||| |||||
Db 1864 TGAACGGCTACCGATGA 1847

RESULT 4
5460961-1/c
Patent No. 5460961

APPLICANT: DEBY, CAROL; PINCEMAIL, JOEL; BOLLEN, ALEX
TITLE OF INVENTION: HUMAN MYELOPEROXIDASE AND ITS

THERAPEUTIC APPLICATION
NUMBER OF SEQUENCES: 13

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/641,678

FILING DATE: 16-JAN-1991
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 460,931
FILING DATE: 14-FEB-1990

SEQ ID NO: 1
LENGTH: 2261

5460961-1

Query Match 68.3%; Score 16.4; DB 6; Length 2261;
Best Local Similarity 94.4%; Pred. No. 41;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 2 tgaactgtgacgatga 19
||||| ||||| ||||| |||||
Db 2016 TGAACGGCTACCGATGA 1999

RESULT 5
US-09-328-111-816
Sequence 816, Application US/09328111

Patent No. 6262333
GENERAL INFORMATION:

APPLICANT: Endege, Wilson O.
APPLICANT: Steinmann, Kathleen E.

APPLICANT: Astle, Jon H.
APPLICANT: Burgess, Christopher C.

APPLICANT: Bushnell, Steven E.
APPLICANT: Carroll III, Eddie

APPLICANT: Catino, Theodore J.
APPLICANT: Dertl, Adnan

APPLICANT: Ford, Donna M.
APPLICANT: Lewis, Marcia E.

APPLICANT: Monahan, John E.
APPLICANT: Schlegel, Robert

TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
PRODUCTS

FILE REFERENCE: CCD-257 (US)
CURRENT APPLICATION NUMBER: US/09/328,111

CURRENT FILING DATE: 1999-06-08

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 6, 2002, 09:48:47 : Search time 256.95 Seconds
(without alignments)
80.077 Million cell updates/sec

Title: US-09-897-042-13

Perfect score: 1 atgactgtgtaccagtgactaa 24

Sequence: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Scoring table: 930621 segs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

N_Geneseq_1101:*

1: /SIDS2/gcgdata/geneseq/geneseq/NA1980.DAT:*

2: /SIDS2/gcgdata/geneseq/geneseq/NA1981.DAT:*

3: /SIDS2/gcgdata/geneseq/geneseq/NA1982.DAT:*

4: /SIDS2/gcgdata/geneseq/geneseq/NA1983.DAT:*

5: /SIDS2/gcgdata/geneseq/geneseq/NA1984.DAT:*

6: /SIDS2/gcgdata/geneseq/geneseq/NA1985.DAT:*

7: /SIDS2/gcgdata/geneseq/geneseq/NA1986.DAT:*

8: /SIDS2/gcgdata/geneseq/geneseq/NA1987.DAT:*

9: /SIDS2/gcgdata/geneseq/geneseq/NA1988.DAT:*

10: /SIDS2/gcgdata/geneseq/geneseq/NA1989.DAT:*

11: /SIDS2/gcgdata/geneseq/geneseq/NA1990.DAT:*

12: /SIDS2/gcgdata/geneseq/geneseq/NA1991.DAT:*

13: /SIDS2/gcgdata/geneseq/geneseq/NA1992.DAT:*

14: /SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT:*

15: /SIDS2/gcgdata/geneseq/geneseq/NA1994.DAT:*

16: /SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT:*

17: /SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT:*

18: /SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT:*

19: /SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT:*

20: /SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT:*

21: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:*

22: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	17.6	73.3	33030	22	AAF29337
2	16.6	69.2	573	21	AAC49803
3	16.6	69.2	823	21	AAC35318
4	16.6	69.2	1452	20	AAK61284
5	16.6	69.2	1452	20	AAK61284
6	16.4	68.3	237	19	AAV11521
7	16.4	68.3	257	22	AAI25101
8	16.4	68.3	257	22	AAI50972
9	16.4	68.3	444	19	AAV11515
10	16.4	68.3	615	19	AAV11514
11	16.4	68.3	702	19	AAV11509

C 12	16.4	68.3	1386	19	AAV11507
C 13	16.4	68.3	2260	11	AAQ03118
C 14	16.4	68.3	3215	18	AAT66437
C 15	16.2	67.5	424	21	AAC31939
C 16	16.2	67.5	516	21	AAC54535
C 17	16.2	67.5	586	21	AAC34988
C 18	16.2	67.5	658	21	AAZ80732
C 19	16.2	67.5	1269	17	AAT15596
C 20	16.2	67.5	1344	18	AAT73169
C 21	16.2	67.5	1349	18	AAT73177
C 22	16.2	67.5	1350	18	AAT73176
C 23	16.2	67.5	1366	18	AAT73174
C 24	16.2	67.5	1374	18	AAT73173
C 25	16.2	67.5	1395	18	AAT73172
C 26	16.2	67.5	1401	18	AAT73167
C 27	16.2	67.5	1412	18	AAT73175
C 28	16.2	67.5	1417	18	AAT73171
C 29	16.2	67.5	1441	18	AAT73170
C 30	16.2	67.5	1967	21	AAC78171
C 31	16.2	67.5	3239	22	AAD05543
C 32	16.2	67.5	3265	22	AAD08193
C 33	16.2	67.5	3265	22	AAD05545
C 34	16.2	67.5	3296	22	AAD12580
C 35	16.2	67.5	3320	22	AAS23196
C 36	16.2	67.5	3339	22	AAD05503
C 37	16.2	67.5	3339	22	AAD05544
C 38	16.2	67.5	3349	21	AAAI6628
C 39	16.2	67.5	3362	22	AAF93802
C 40	16.2	67.5	3404	17	AAI55595
C 41	16.2	67.5	3503	19	AAV07164
C 42	16.2	66.7	146	22	AAI51261
C 43	16.2	66.7	285	21	AAC94537
C 44	16.2	66.7	317	20	AAV88731
C 45	16.2	66.7	360	14	AAO61040

ALIGNMENTS

RESULT 1	
AAF29337	standard: DNA: 33030 BP.
XX	
XX	AAF29337:
XX	
DT	20-APR-2001 (first entry)
XX	
DE	Atopy related gene sequence SEQ ID 1.
XX	
KW	Atopy: human; allergic disease; atopic dermatitis; asthma; ds.
XX	
OS	Homo sapiens.
XX	
PN	WO200104302-A1.
XX	
PD	18-JAN-2001.
XX	
PF	10-JUL-2000; 2000WO-JP04599.
XX	
PR	09-JUL-1999; 99JP-0196517.
XX	
PA	(MITU) MITSUBISHI CHEM CORP.
XX	
XX	Adra CM, Shitake T, Hopkin JM.
DR	WPI: 2001-138350/14.
XX	
PT	New atopy gene with domain containing H7M4 gene in eleventh chromosome
PT	q13.1 domain identified and isolated by PCR amplification, for
PT	diagnosis and drug development e.g. for allergic diseases and asthma
PS	Claim 1; Page 18-31; 47pp; Japanese.
XX	

Recombinant MPO DN
HindIII/HpaI frag
Myeloperoxidase co
Human secreted pro
Arabidopsis thalia
Arabidopsis thalia
Human colon cancer
Thermostable cellu
Pseudomonas putida
Pseudomonas putida
Pseudomonas putida
Pseudomonas putida
Pseudomonas putida
Pseudomonas putida
Pseudomonas putida
Pseudomonas putida
Human cancer assoc
Human secreted pro
Human secreted pro
Human secreted pro
Human protein havi
DNA encoding novel
Human secreted pro
Human secreted pro
Human secreted pro
Human cDNA encodin
Thermostable cellu
Thermomosporea fu
Probe #19947 used
Cat flea hindgut a
EST clone H0253.
Human brain Expres

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 6, 2002, 09:45:57 ; Search time 2642.04 Seconds
(without alignments)
97.613 Million cell updates/sec

Title: US-09-897-042-13

Perfect score: 24
Sequence: 1 atgaactgtgacgcgatgaactaa 24

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

EST: *
1: em_estfun: *
2: em_esthum: *
3: em_estlin: *
4: em_estom: *
5: em_estpl: *
6: em_estba: *
7: em_estro: *
8: em_estov: *
9: em_hlc: *
10: qb_estl: *
11: qb_est2: *
12: qb_hlc: *
13: qb_gss: *
14: em_gss_fun: *
15: em_gss_hum: *
16: em_gss_inv: *
17: em_gss_pln: *
18: em_gss_pro: *
19: em_gss_rtd: *
20: em_gss_vrt: *
21: em_gss_other: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	18.4	76.7	577	13	AQ453713 HS_5155_A
2	18.2	75.8	181	10	AA585156 ATH637 HT
3	18.2	75.8	228	10	BB258711 BB258711
4	18.2	75.8	495	10	AU221054 AU221054
5	18.2	75.8	648	10	AI054495 coau0001C
6	18.2	75.8	763	13	CNS0255R
7	17.8	74.2	502	10	AL181656 Tetradon
8	17.8	74.2	1039	13	AW017100 EST272522
9	17.6	73.3	328	10	AL394021 T7 end of
10	17.6	73.3	370	10	BB120380 BB120380
11	17.6	73.3	427	13	AV622113 AV622113
12	17.6	73.3	500	13	AZ144229 SP_0020_B
13	17.6	73.3	500	13	AZ173900 SP_0125_B

C 13	17.6	73.3	533	11	BF349591
C 14	17.6	73.3	560	13	AZ794406
C 15	17.6	73.3	562	13	AZ154963 SP_0040_A
C 16	17.6	73.3	593	13	BH022435 GH_MBB000
C 17	17.6	73.3	677	11	BG856621 1024047C0
C 18	17.6	73.3	682	11	BG848783
C 19	17.6	73.3	764	11	BG848784
C 20	17.6	73.3	809	11	BG784924
C 21	17.6	73.3	921	13	AZ201725 SP_0055_A
C 22	17.6	73.3	946	13	CNS010XD
C 23	17.6	73.3	1045	13	CNS060YH
C 24	17.4	72.5	338	10	AA932931
C 25	17.4	72.5	393	10	AA905887
C 26	17.4	72.5	400	13	AO505888
C 27	17.4	72.5	428	11	T89283
C 28	17.4	72.5	483	11	BG466615
C 29	17.4	72.5	506	10	AM654839
C 30	17.4	72.5	523	10	AM444944
C 31	17.4	72.5	677	10	AV654637
C 32	17.2	71.7	303	10	BB202920
C 33	17.2	71.7	407	13	AO033034
C 34	17.2	71.7	424	11	BG042989
C 35	17.2	71.7	467	10	BE111384
C 36	17.2	71.7	471	10	BE359101
C 37	17.2	71.7	479	11	BI075714
C 38	17.2	71.7	480	10	AU203508
C 39	17.2	71.7	518	10	AA409213
C 40	17.2	71.7	557	10	AJ280888
C 41	17.2	71.7	573	11	BG466085
C 42	17.2	71.7	629	10	AM828984
C 43	17.2	71.7	641	10	AW353303
C 44	17.2	71.7	663	11	BG063351
C 45	17.2	71.7	709	13	AZ561191

ALIGNMENTS

RESULT 1
LOCUS AQ453713 577 bp DNA GSS 21-Apr-1999
DEFINITION HS_5155_A2_E05.SP6E RPT-11 Human Male BAC Library Homo sapiens
ACCESSION AQ453713
VERSION AQ453713.1 GI:4594883
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 577)
AUTHORS Mahairas G.G., Wallace J.C., Smith R., Swartzell S., Holzman T.,
Keller A., Shaker R., Furlong J., Young J., Zhao S., Adams M.D. and
Hood L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPT-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering/bac.htm>)
or from Research Genetics (info@resgen.com). BAC end Web Server:
<http://www.htsc.washington.edu>
Plate: 731 row: I column: 10
Seq primer: SP6

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 6, 2002, 09:47:47 : Search time 1586.23 Seconds
(without alignments)
249.606 Million cell updates/sec

Title: US-09-897-042-13

Perfect score: 1 atgaactgtgtaacgaactaa 24

Sequence: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Scoring table: 1472140 segs, 8248589755 residues

Searched: Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl :
1: gb_da :
2: gb_htg :
3: gb_in :
4: gb_om :
5: gb_ov :
6: gb_pat :
7: gb_ph :
8: gb_pl :
9: gb_pr :
10: gb_ro :
11: gb_sts :
12: gb_sy :
13: gb_un :
14: gb_vl :
15: em_ba :
16: em_fun :
17: em_hum :
18: em_in :
19: em_om :
20: em_or :
21: em_ov :
22: em_pat :
23: em_ph :
24: em_pl :
25: em_ro :
26: em_sts :
27: em_sy :
28: em_un :
29: em_vl :
30: em_htgo_hum :
31: em_htgo_inv :
32: em_htgo_rod :
33: em_htg_hum :
34: em_htg_inv :
35: em_htg_rod :
36: em_htg_other :

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	19.8	82.5	57329	9	AL391649	AL391649 Human DNA
C 2	19.8	82.5	162149	2	AC023574	AC023574 Homo sapi
C 3	19.2	80.0	170343	2	AC015939	AC015939 Mus muscu
C 4	19.2	80.0	174162	2	AC024611	AC024611 Mus muscu
C 5	19.2	80.0	241880	10	AC027285	AC027285 Mus muscu
C 6	19.2	80.0	243267	2	AC011407	AC011407 Homo sapi
C 7	18.2	75.8	32962	3	CELR05H11	U00056 Caenorhabdi
C 8	18.2	75.8	58949	3	CELR07E12	U00054 Caenorhabdi
C 9	18.2	75.8	166536	2	AC018465	AC018465 Homo sapi
C 10	17.8	74.2	29482	2	CEW05B5	282071 Caenorhabdi
C 11	17.8	74.2	89004	3	CEY65H2	Z58877 Caenorhabdi
C 12	17.8	74.2	102433	8	AY013246	AY013246 Hordeum v
C 13	17.8	74.2	145576	8	AP002541	AP002541 Oryza sat
C 14	17.8	74.2	147857	8	AP002747	AP002747 Oryza sat
C 15	17.6	73.3	2540	8	SPRAP1	X57078 S.pombe pap
C 16	17.6	73.3	4993	1	CAU17110	U17110 Drosophilid
C 17	17.6	73.3	11949	1	U32808	U32808 Haemophilus
C 18	17.6	73.3	13403	1	AE007768	AE007768 Clostridi
C 19	17.6	73.3	17487	3	DMC171E4	AL021726 Drosophil
C 20	17.6	73.3	19469	8	SPAC1783	AL136499 S.pombe c
C 21	17.6	73.3	21693	2	AC017390	AC017390 Drosophil
C 22	17.6	73.3	21838	1	PDCRAFOEP	L52093 Pedicoccus
C 23	17.6	73.3	21839	1	PEPUSREOP	Z32771 P.pentlosace
C 24	17.6	73.3	31144	2	AC087398	AC087398 Homo sapi
C 25	17.6	73.3	51858	9	AF097649	AF097649 Homo sapi
C 26	17.6	73.3	110000	2	AC092449	AC092449 Homo sapi
C 27	17.6	73.3	112363	2	AC087425	AC087425 Oryza sat
C 28	17.6	73.3	115861	9	AC006971	AC006971 Homo sapi
C 29	17.6	73.3	123209	9	AL158195	AL158195 Human DNA
C 30	17.6	73.3	143107	9	AC005949	AC005949 Homo sapi
C 31	17.6	73.3	145012	8	AC087551	AC087551 Oryza sat
C 32	17.6	73.3	149684	9	AL355485	AL355485 Human DNA
C 33	17.6	73.3	154065	9	AC005948	AC005948 Homo sapi
C 34	17.6	73.3	156292	9	CNS05TD6	AL356016 Human chr
C 35	17.6	73.3	193430	2	AC011709	AC011709 Homo sapi
C 36	17.6	73.3	198127	2	AL512510	AL512510 Homo sapi
C 37	17.6	73.3	210613	2	AC044907	AC044907 Homo sapi
C 38	17.6	73.3	296497	3	AE003421	AE003421 Drosophil
C 39	17.4	72.5	885	9	AF143325	AF143325 Homo sapi
C 40	17.4	72.5	40118	9	HS063113	U63313 Human cosm
C 41	17.4	72.5	41935	9	HS061B11	Z73913 Human DNA s
C 42	17.4	72.5	64356	9	HS512B11	AL031058 Human DNA
C 43	17.4	72.5	71990	2	AL590044	AL590044 Homo sapi
C 44	17.4	72.5	148434	2	AL133543	AL133543 Homo sapi
C 45	17.4	72.5	169949	9	AC084358	AC084358 Homo sapi

ALIGNMENTS

RESULT 1
AL391649/c 57329 bp DNA PRI 12-JAN-2001
LOCUS Human DNA sequence from clone RP11-202E10 on chromosome 9 Contains
DEFINITION GSSs and STSs, complete sequence.
ACCESSION AL391649
VERSION AL391649.8 GI:11878445
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
Mammalia: Eutheria: Primates: Catarrhini: Homiidae: Homo.
REFERENCE 1 (bases 1 to 57329)
AUTHORS Sehara,H.
TITLE Direct Submission
JOURNAL Submitted (04-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone
requests: clonerequests@sanger.ac.uk
On Dec 18, 2000 this sequence version replaced gi:11876078.
During sequence assembly data is compared from overlapping clones.

COMMENT

Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated repeat sequence elements. Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr9>

Rp11-202E10 is from the library RPC1-11.1 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: BPAC3.6

IMPORTANT: This sequence is not the entire insert of clone Rp11-202E10. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.

The true left end of clone Rp11-477G9 is at 57230 in this sequence. The true right end of clone Rp11-336I6 is at 100 in this sequence.

FEATURES

source

1..57329

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="9"

/clone="Rp11-202E10"

/clone_lib="RPC1-11.1"

/complement(1..100)

/note="match: STS: Em:G50491"

match: GSS: Em:AQ046842"

110..609

/note="match: GSS: Em:AQ834379"

1506..1774

/note="L2 repeat: matches 2263..2542 of consensus"

2279..2402

/note="FLAM_A repeat: matches 2..127 of consensus"

2409..2705

/note="AluY repeat: matches 1..297 of consensus"

2723..2864

/note="BC200 repeat: matches 1..142 of consensus"

3300..3830

/note="match: GSS: Em:AQ666909"

/complement(4552..4767)

/note="match: STS: Em:G04418"

4886..5133

/note="AluSg repeat: matches 3..290 of consensus"

5633..5708

/note="38 copies 2 mer 1t 65% conserved"

5714..5910

/note="L1PB3 repeat: matches 5951..6150 of consensus"

5911..6273

/note="MSTRA repeat: matches 1..426 of consensus"

6274..7925

/note="MSTRA-internal repeat: matches 1..1651 of consensus"

7926..8304

/note="MSTRA repeat: matches 1..426 of consensus"

8305..8814

/note="L1PB3 repeat: matches 5455..5951 of consensus"

9190..10603

/note="L1MA2 repeat: matches 4903..6308 of consensus"

10604..11035

/note="L1PB2 repeat: matches 5712..6155 of consensus"

11036..11809

/note="L1MA2 repeat: matches 4128..4903 of consensus"

14222..14608

/note="MIR repeat: matches 59..172 of consensus"

repeat_region

/note="HAI1 repeat: matches 397..791 of consensus"

repeat_region

/note="L1MC4 repeat: matches 7860..7956 of consensus"

misc_feature

/note="AluSc repeat: matches 1..308 of consensus"

repeat_region

/complement(15008..15360)

repeat_region

/note="match: GSS: Em:B37456"

repeat_region

/note="match: GSS: Em:B37456"

repeat_region

/note="AluSx repeat: matches 1..312 of consensus"

repeat_region

/note="AluSx repeat: matches 1..312 of consensus"

repeat_region

/note="L1MD repeat: matches 973..1658 of consensus"

repeat_region

/note="AluDb repeat: matches 1..286 of consensus"

repeat_region

/note="AluDb repeat: matches 1..286 of consensus"

repeat_region

/note="L1MA9 repeat: matches 4078..4330 of consensus"

repeat_region

/note="L1MA9 repeat: matches 4078..4330 of consensus"

repeat_region

/note="L1MA9 repeat: matches 5782..6275 of consensus"

repeat_region

/note="L1MA9 repeat: matches 5782..6275 of consensus"

repeat_region

/note="L1MA9 repeat: matches 5272..5789 of consensus"

repeat_region

/note="L1MA9 repeat: matches 5272..5789 of consensus"

repeat_region

/note="L1MA9 repeat: matches 2624..2750 of consensus"

repeat_region

/note="L1MA9 repeat: matches 2624..2750 of consensus"

repeat_region

/note="AluY repeat: matches 3..308 of consensus"

repeat_region

/note="AluY repeat: matches 3..308 of consensus"

repeat_region

/note="L1MA9 repeat: matches 1..303 of consensus"

repeat_region

/note="L1MA9 repeat: matches 1..303 of consensus"

repeat_region

/note="L1MA9 repeat: matches 1..303 of consensus"

repeat_region

/note="L1MA9 repeat: matches 1..303 of consensus"

repeat_region

/note="L1MA9 repeat: matches 1..303 of consensus"

repeat_region

/note="L1MA9 repeat: matches 1..303 of consensus"

repeat_region

/note="L1MA9 repeat: matches 1..303 of consensus"

repeat_region

/note="L1MA9 repeat: matches 1..303 of consensus"

repeat_region

/note="L1MA9 repeat: matches 1..303 of consensus"

repeat_region

/note="L1MA9 repeat: matches 1..303 of consensus"

repeat_region

/note="L1MA9 repeat: matches 1..303 of consensus"

repeat_region

/note="L1MA9 repeat: matches 1..303 of consensus"

repeat_region

/note="L1MA9 repeat: matches 1..303 of consensus"

repeat_region

/note="L1MA9 repeat: matches 1..303 of consensus"

repeat_region

/note="L1MA9 repeat: matches 1..303 of consensus"

repeat_region

/note="L1MA9 repeat: matches 1..303 of consensus"

repeat_region

/note="L1MA9 repeat: matches 1..303 of consensus"

repeat_region

/note="L1MA9 repeat: matches 1..303 of consensus"

repeat_region

/note="L1MA9 repeat: matches 1..303 of consensus"

repeat_region

/note="L1MA9 repeat: matches 1..303 of consensus"

repeat_region

/note="L1MA9 repeat: matches 1..303 of consensus"